Chapter 14
Segregation and Sentiment: Estimating Refugee Segregation and Its Effects Using Digital Trace Data

Neal Marquez, Kiran Garimella, Ott Toomet, Ingmar G. Weber and Emilio Zagheni

Abstract In light of the ongoing events of the Syrian Civil War, many governments have shifted the focus of their hospitality efforts from providing temporary shelter to sustaining this new long-term population. In Turkey, a heightened focus has been placed on the encouragement of integration of Syrian refugees into Turkish culture, through the dismantling of Syrian refugee-only schools in Turkey and attempts to grant refugees permanent citizenship, among other strategies. Most of the existing literature on the integration and assimilation of Syrian refugees in Turkey has taken the form of surveys assessing the degree to which Syrian refugees feel they are part of Turkish culture and the way Turkish natives view the refugee population. Our analysis leverages call detail record data, made available by the Data for Refugees (D4R) Challenge, to assess how communication and segregation vary between Turkish natives and Syrian refugees over time and space. In addition, we test how communication and segregation vary with measures of hostility from Turkish natives using data from the social media platform Twitter. We find that measures of segregation vary significantly over time and space. We also find that measures of intergroup communication positively correlate with measures of public sentiment toward refugees. Attempts to address the concerns of Turkish natives in order to minimize the traction of online hate movements may help to improve the integration process.

N. Marquez (✉) · E. Zagheni
University of Washington, Washington, DC, USA
e-mail: nmarquez@uw.edu

Max Planck Institute for Demographic Research, Rostock, Germany

K. Garimella
Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

O. Toomet
University of Washington, Washington, DC, USA

I. G. Weber
Qatar Computing Research Institute, Doha, Qatar

A. A. Salah et al. (eds.), *Guide to Mobile Data Analytics in Refugee Scenarios*, https://doi.org/10.1007/978-3-030-12554-7_14

14.1 Introduction

In the spring of 2011, at the beginning of the Syrian Civil War, Syrians began to find themselves displaced by the armed conflicts between the Syrian Arab Republic and numerous other forces who sought to challenge the authority of the government in the wake of perceived injustices committed by the regime led by Bashar al-Assad [1]. During this time, Turkey had an open door policy with Syria and assured that those migrating in would be able to stay until Syria was once again safe for return [2]. By later that year, it was apparent that extensive measures would need to be taken to accommodate the growing number of refugees. During the first years of the Syrian conflict, it was unclear how long the crisis would last and require refugees to seek asylum, in Turkey and other locations. Initial measures addressed short-term issues by setting up temporary schools, camps, and healthcare facilities [2]. By 2015, however, it became clear that the conflict was not to conclude in the near future and the flow of refugees into Turkey continued, reaching over 2.5 million Syrians in Turkey by the end of the year.[1]

The strategy of the Turkish government shifted from short- to long-term plans, as policies were developed to ease the transition of Syrians into Turkish life. A new, worldwide visibility of the plight of Syrian refugees allowed Turkey to coax greater action from the international community to share a portion of the economic and resource burden created by housing refugees. Though other European countries have stepped up their contributions to the crisis by way of accepting more refugees and offering Turkey financial compensation [3], Turkey has by far the largest Syrian refugee population to date, more than 3.5 million as of August 2018,[2] and continues to struggle with integrating the population. The difficulty of integrating refugees into Turkish culture is a battle that has two fronts, as the government not only looks to facilitate a smooth transition for refugees but also to ease the concerns of Turkish natives, who fear the extended stay of Syrian refugees may come at the expense of their desired lifestyle [4].

Geographic segregation and social isolation can exacerbate the differences between these two groups by limiting the amount of cultural overlap they experience. To date almost no measures of segregation of Syrian refugees and Turkish natives are available. The rapid increase in the number of refugees in Turkey in the past few years has made it difficult for traditional methods of data collection to capture this phenomenon.

This analysis leverages call detail record (CDR) data, made available by the Data for Refugees (D4R) Challenge, to assess how communication and segregation between Turkish natives and Syrian refugees differ over time and space. Using CDR data, we create metrics of geographic activity space and residential dissimilarity as measures of segregation. We also calculate spatial–temporal measures of the probability of refugees contacting Turkish citizens through phone calls and texts, as a measure of group isolation. Finally, we examine how communication between the

[1] https://data2.unhcr.org/en/situations/syria/location/113.

[2] https://data2.unhcr.org/en/situations/syria/location/113.

two groups is altered by differing levels of segregation as well as changes of expressed opinions from Turkish citizens toward Syrian refugees by leveraging discussion by Turkish natives of Syrian refugees on the social media platform Twitter.

14.2 Background

14.2.1 Segregation

Segregation has long been seen as a mechanism that isolates individuals from accessing greater opportunities if their isolated enclave is poor in group resources [5]. In addition, greater isolation of communities has been linked to increased xenophobic attitudes toward minority migrant groups in the Global South [6]. Previous policy research has advocated for working toward greater cohesion between groups in the form of public education campaigns as a way of combating negative opinions toward these minority groups [7]. The extent to which segregation between native populations and refugees is an issue in Turkey is not yet well understood.

To date, no studies have systematically or comprehensively quantified the level to which segregation exists between Syrian refugees and Turkish natives. We use the word segregation here simply to mean the separation of two or more groups of people, in our case Syrian refugees and individuals native to Turkey. Furthermore, the study of drivers and consequences of segregation have been studied only in limited contexts, such as economic consequences. For example, in a recent publication by Balkan et al. [8], the authors found that increases in the refugee population led to increased rent costs in higher end properties, which is seen as evidence for increased value of housing that is geographically segregated from refugee populations [8]. Additionally, İcduygu et al. [9] found that integration efforts made by Syrian refugees to participate in the legal labor force were thwarted by difficulties to obtain visas, thus limiting chances to integrate socially and culturally [9].

14.2.2 Turkish Attitudes Toward Refugees

How segregation ties in with attitudes from Turkish citizens is at the moment unclear. This is due to a number of factors that have restricted measurement of segregation even as the Syrian refugee population growth has slowed down in 2019. Because the level of segregation between the two populations is not well known in Turkey, it is difficult to discuss the potential effect that it has on Turkish citizens' opinions of refugees, if any at all. We do know, however, that sentiment towards refugees has recently been negatively trending. While early studies showed a more neutral stance on the Syrian refugee population, recent studies show strong negative attitudes [2]. In the Syrian Barometer Study 2017, Erdoğan found that over 80% of Turkish survey

respondents claimed that the Syrian and Turkish cultures do not overlap at all [10]. In addition, several studies have found that some populations who have experienced large Syrian refugee intake have taken to social media platforms to voice their dissatisfaction with the presence and government handling of Syrian refugees [11, 12].

Social media platforms, such as Twitter, offer a way to study how populations react to events without the time or expense requirements of conducting a survey. While Twitter is known to not have a representative population of users, studies have found that text analysis in the form of sentiment extraction can provide reliable predictions for population-wide events [13]. Additionally, researchers have been able to track changes in attitudes toward minority groups in response to policy announcements [14].

More recent studies have begun to directly examine how citizens talk about refugees in their home country on various social media platforms. A content analysis of tweets—posts from the social media platform Twitter—about Syrian refugees across Europe found that when users attack refugees, they often do so by attacking the character of male refugees, labeling them either as cowards or terrorists [12]. Another report within Turkey found that several anti-Syrian hashtags had gained traction in 2017, undermining efforts to foster greater cohesion between refugees and citizens [15]. The events co-occurred with a threefold increase in intergroup violence between 2016 and 2017, lending evidence that events on Twitter may in fact well represent attitudes of the greater population despite Twitter only having a 15% penetration rate in Turkey [16].

Our analysis tests how segregation, both geographic and social, varies over space and time between Syrian refugees and Turkish natives using CDR data. Using this information, we will be able to make better informed decisions regarding the way that refugees have integrated into the Turkish population deferentially within the country. We can do this by examining both residential and activity space dissimilarity as measures of geographic segregation. Furthermore, we can quantify social isolation by assessing the kind of persons that refugees call, either fellow refugees or Turkish citizens.

Lastly, using Twitter data that contain subjects related to refugees, we will examine how variation in the sentiment of tweets alters with changes in refugee-citizen segregation over space and time. Twitter has been home to many discussions related to Syrian refugees, both positive and negative, especially within Turkey [12]. By analyzing how fluctuations in discussion co-occur with changes in segregation, we may obtain a better understanding of how the two social process influence one another.

14.3 Data

The analysis utilizes call detail records (CDR) from the Turkish mobile network carrier Turk Telecom (TT), a member of the group TTG, as part of the Data for Refugees in Turkey (D4R) challenge [17]. The goal of the challenge is to give researchers access to privately owned data from TTG that has user details removed for anonymity, such

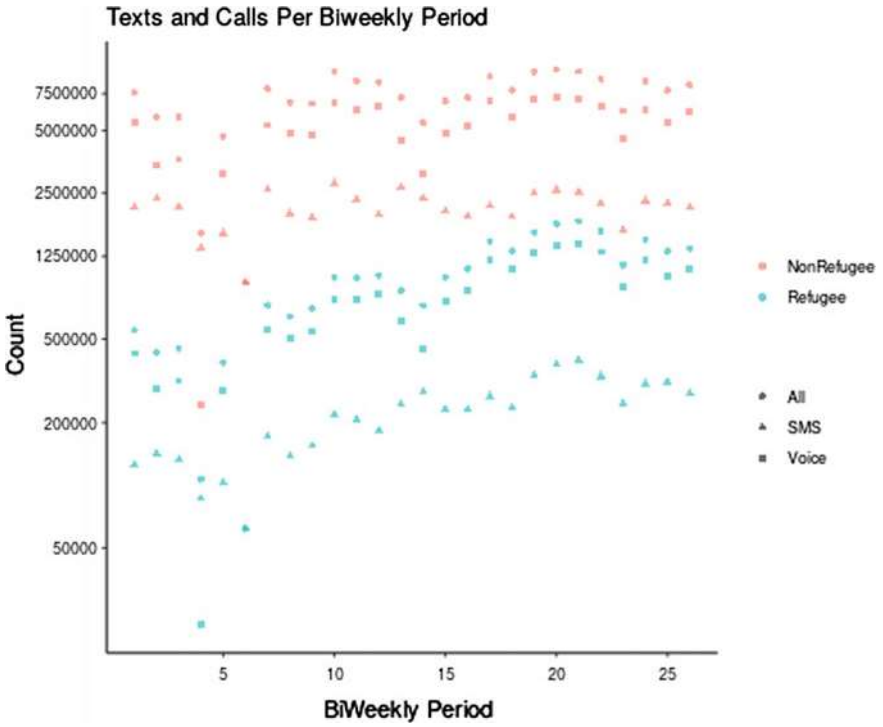

Fig. 14.1 Number of Texts and Calls present in each biweekly dataset broken down by ID type of the TTG user, either Registered Refugee or non-Refugee. Values are shown on a log scale

as names and telephone numbers. The time stamp and the location of the call are available in the data set. Each call also has a randomized ID assigned to it, which indicates a unique user and whether TTG has the individual recorded as a refugee or not. This classification does not perfectly identify refugees and should be seen as an imperfect measure [17]. The specific data set that we utilize in our analysis tracks users for 2 weeks at a time with an undisclosed portion of their calls and text messages both sent and received provided. The CDRs provide time stamp data, to the hour, and the cell phone tower that was pinged for the particular record. The records consist of 212,364,027 unique records from 5,006,222 and 1,082,603 unique non-refugee and refugee users, respectively. The call records span 26 two-week segments from January 1, 2017 to December 31, 2017 with the number of calls and users being unequally distributed across time (Fig. 14.1).

Individuals were over-sampled for areas that had relatively high refugee populations, such as border provinces and the major metropolitan areas of Istanbul and Izmir [18]. Each record in the data is given a tower ID which can be linked via a database with towers and their corresponding latitude and longitude. Any tower with a location outside of Turkey's administrative bounds was removed from the dataset. To verify that we can capture mobility of individuals to an adequate level,

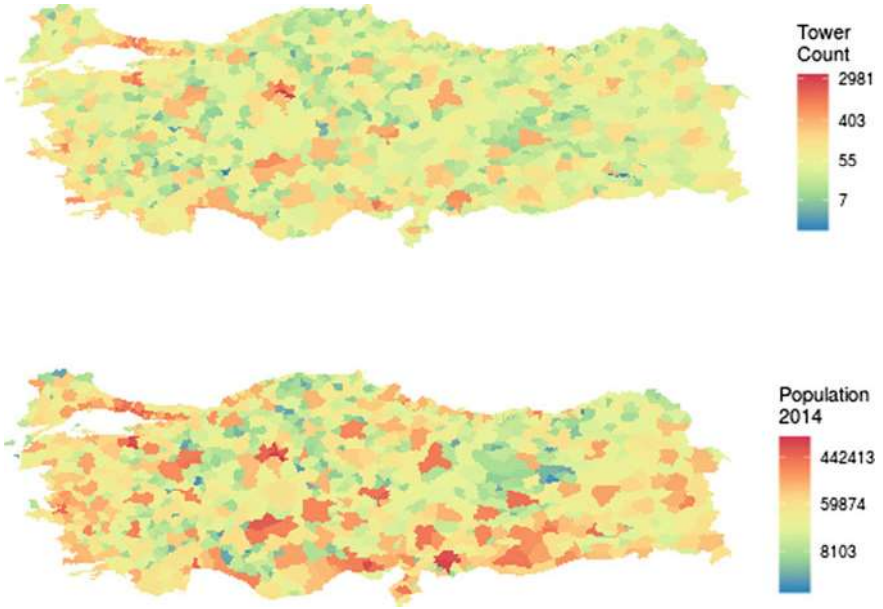

Fig. 14.2 Population counts (lower panel) & cell phone tower users (upper panel) geographic distributions at the district level

we analyzed the degree to which district-level (administrative level 2) population size correlated with the number of cell phone towers in an area. In a log-l-og linear model the tower count explained 81.7% of the variation in the 2014 population, taken from the 2014 Turkey national census, at the district level. The areas that had the most discrepancy between the number of cellular towers and the population count can be seen in Fig. 14.2. Refugee status of the other individual participating in the phone call is also provided in the dataset.

To estimate changing attitudes over time and space in Turkey, we pulled Twitter data from 2011 to 2017 from the Twitter Stream that matched several topics related to Syrian refugees (see Appendix). The Twitter Stream is an ongoing project from the Internet Archive Team that consistently collects a 1% stream of all Twitter data produced.[3] While the Twitter API only allows users to collect data that has recently been created, this archive allows us to search trends that overlap with our CDR record dataset. Tweets were only considered for our statistical analysis if they were from 2017. We further restricted our analysis to include only tweets from Turkish language users, users that specified their location to be within Turkey, or tweets that could be geolocated within Turkey. Individual tweets could be geolocated either by providing the exact coordinates of the location of the tweet, i.e., "Tweet with a location" option, or by designating a "place" from a prespecified list provided by Twitter which contains geographic coordinates. If these coordinates fall within the

[3] https://archive.org/details/twitterstream.

administrative boundaries of Turkey, the tweets are kept. Users could be identified as being from Turkey based on their user-specific location string. To geotag this string, we use the Open Street Maps API and select the location coordinates with the highest match to determine if the user is located within Turkey. This filtering process left us with 65,778 tweets for our analysis.

Several other variables were collected for modeling purposes. Population data at the province level was taken from the 2014 Turkish census. More recent population estimates now exist, however, were not readily available to the authors at the time of analysis. Land use data was collected from CORINE Land Cover surveys 2006–2012 to calculate the percent human created land coverage, a proxy measure for urban space [19]. These data were then population weighted using population rasters created by satellite imagery from the gridded population of the world v4 [20].

14.4 Methods

To calculate residential and activity space dissimilarity for a district, we created subunits within each district by way of Voronoi tessellation from the cell phone towers within the district. Voronoi tessellation creates areal units which define a two-dimensional space that is the least distance from a particular point, in our case a cell phone tower [21]. If many towers exist in a district, then the areas that are created are relatively granular, given that the towers are evenly spaced. Using Voronoi cells as subdivisions of districts, we calculate an activity space dissimilarity index for each district. While traditional residential dissimilarity indexes measure differences from the perspective that individuals are situated in a single location, activity space dissimilarity measures the probability of remaining isolated from another group or 1—"potential to encounter" as defined in Wong et al. [22]. Activity space dissimilarity scores were calculated for each district for each week of the analysis using the formula in Eq. 14.2 where i is a Voronoi cell, j is an individual, p_{ij} is the percentage of time individual j is in Voronoi cell i, A is the refugee population size, and B is the non-refugee population size. In addition, we calculate residential dissimilarity by taking the modal call location of an individual between the hours of 9 p.m. and 6 a.m. and calculate a traditional dissimilarity index, using the modal location as the place of residence, with the formula in Eq. 14.1. For residential dissimilarity, we only calculated one score per district rather than weekly scores because the values did not change significantly over time, which is to be expected as residential segregation is slow to change.

$$\text{Residential Dissimilarity Score} = \frac{1}{2}\sum_{i}^{N}\left|\frac{a_i}{A} - \frac{b_i}{B}\right| \tag{14.1}$$

$$\text{Activity Space Dissimilarity Score} = \frac{1}{2}\sum_{i}^{N}\left|\frac{\sum_j^A p_{ij}}{A} - \frac{\sum_j^B p_{ij}}{B}\right| \tag{14.2}$$

To test whether the dissimilarity values were different than expected for a district given the number of Voronoi cells and number of refugee and non-refugee calls, we randomized the caller type for each record 1000 times and recalculated dissimilarity scores from the simulated distribution. This procedure is often referred to as a permutation test. Z-scores were then calculated for the district's observed dissimilarity score against the simulated values. Uncertainty for our measures of dissimilarity was calculated by bootstrapping, where individuals were sampled with the replacement for each unit of analysis, district for residential dissimilarity and district-week for activity space dissimilarity.

For each district, we also compiled a connectivity score of refugees to non-refugees as a measure of intercommunication between the two groups. The percentage of calls going from refugees to non-refugees was calculated for each district. We excluded records from non-refugees to refugees because of the small sample size they represented in the data, less than 0.1%.

Tweets were analyzed using a Turkish-translated version of the AFINN, a common sentiment analysis tool with words valence rated on a scale from −5 to 5. Each tweet is rated by the sum of individual word scores. Though this process only allows us to attribute sentiment on a word- by-word basis, it has been extensively tested [23] and is more easily translated into other languages than other sentiment tools. To match Twitter sentiment with CDRs we aggregated sentiment by week and calculated the average weekly sentiment from the Turkish tweets with Syrian related content (Fig. 14.3).

14.4.1 Statistical Model 1: Drivers of Intergroup Calls

To test the relationship between Twitter sentiment and intergroup connectivity, we run a series of logistic regressions, where each outgoing call made by a refugee is the response variable. The outcome is 0 if the call/text was made to a fellow refugee or 1 if made to a non-refugee, with a total of 10,235,988 records. Call records were connected with covariates by their district of call location (for population size, urban area coverage), the biweekly time period that they occurred (for Twitter sentiment), or the combination of the two (for activity space dissimilarity index). We tested a number of covariate combinations to test the robustness of the relationships between covariates and the outcome. To account for the bias in the data from repeated calls from a single user, we ran a mixed effects model with a random intercept on individual. Equation 14.3 shows the structure of the model where i represents an individual, j represents a particular call that was made, β is a vector of beta coefficients, X_{ij} is a vector of coefficients for individual i call j for the particular time and location that

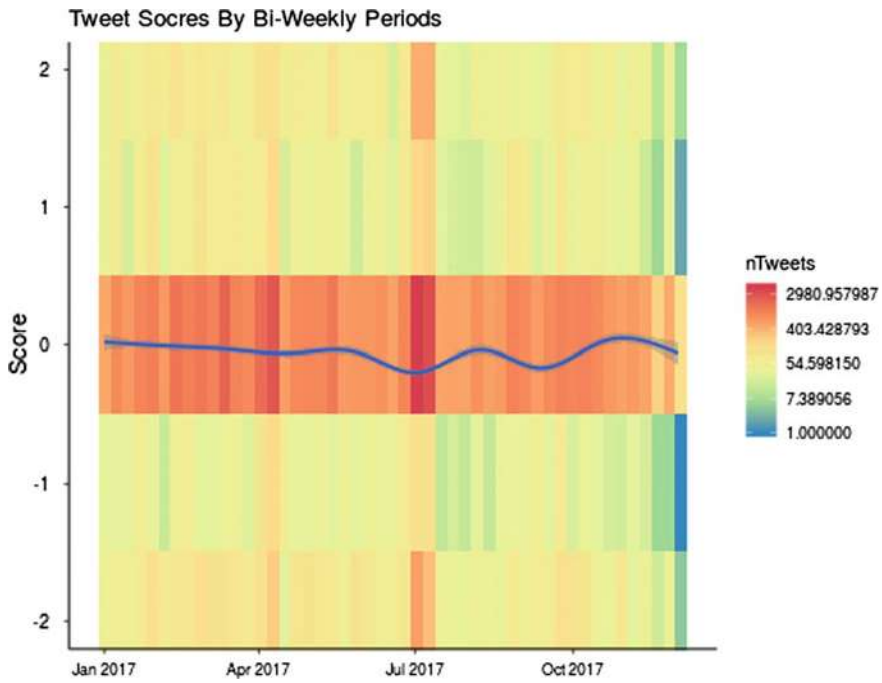

Fig. 14.3 Weekly Sentiment Heat Map with Loess Smoothed Scores. Each rectangular bin is a week(x-axis)–tweet score(y-axis) combination, where the hue indicates the number of tweets in a week that had a particular sentiment score

the call took place, and ζ_i is the individual-level random effect. We did not adjust for spatial autocorrelation as our outcome of interest did not show evidence for it.

$$\begin{aligned} y_{ij} &\sim \text{Binomial}(\hat{p}_{ij}) \\ \hat{p}_{ij} &= \text{logit}(\beta \bullet X_{ij} + \zeta_i) \\ \zeta_i &\sim \mathcal{N}(0, \sigma) \end{aligned} \tag{14.3}$$

14.4.2 Statistical Model 2: Geographic Sentiments

We also tested the ability to predict the sentiment (both positive or negative as well as score) of a tweet as a function of the above mentioned covariates linked by location of the tweet at the province level. Geocoded tweets left us with a considerably smaller sample size from the original dataset, as only 53,793 tweets were from 2017 forward and could be reliably geocoded to a specific province within Turkey. All model covariates were included at the province level and were time invariant.

14.5 Results

Our analysis of spatial overlap found a significant difference between the observed values of activity space dissimilarity and their expected values. Of the 970 districts in our analysis, around 75% had observed values that were more than 4 standard deviations away from their simulated permutation distribution. Of the major metropolitan areas, Ankara had the highest average observed values of dissimilarity, while Istanbul had the lowest, though district-level variance was twice as high in Ankara (Fig. 14.4).

Using bootstrapped estimates of the uncertainty of our calculations for activity space dissimilarity, we found that there were significant differences over time at both the district and province level. We also found that residential dissimilarity was strongly correlated with activity space dissimilarity with a correlation coefficient of $r = 79.96$ ($n = 970$, $p < 0.01$) at the district level and $r = 83.83$ ($n = 81$, $p < 0.01$) at the province level. In line with previous literature, we found that activity space dissimilarity was more often less than residential dissimilarity [24].

Twitter sentiment was also found to change significantly over time but not over locations. Because our province level analysis required that users tweets be geocoded at least to the provincial level, our sample size was dramatically reduced when exam-

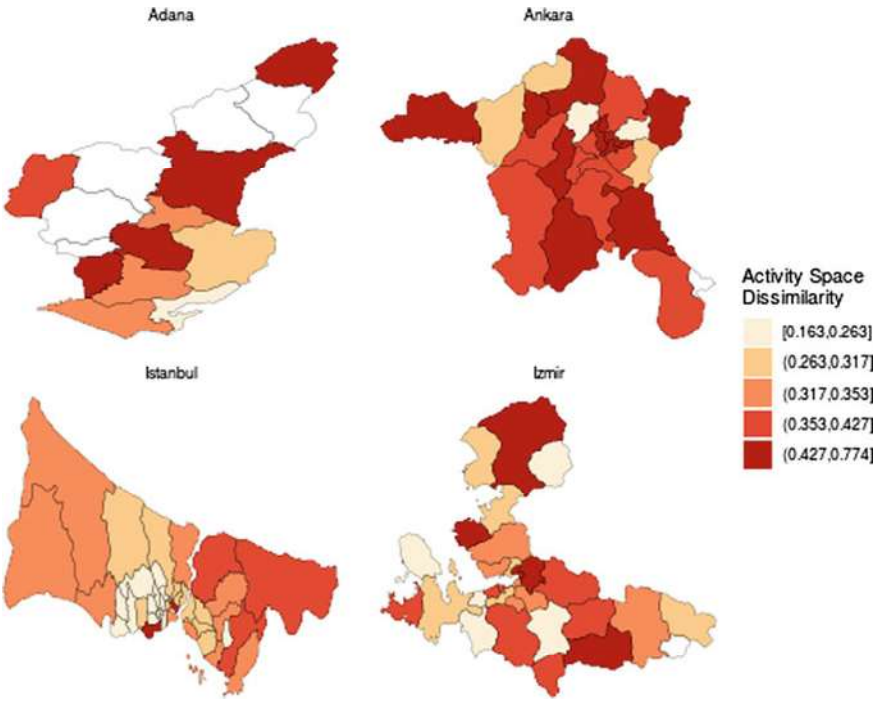

Fig. 14.4 Activity Space Dissimilarity Scores for selected provinces. Results with observed dissimilarity less than 4 standard deviations away from mean of permutation tests are whited out

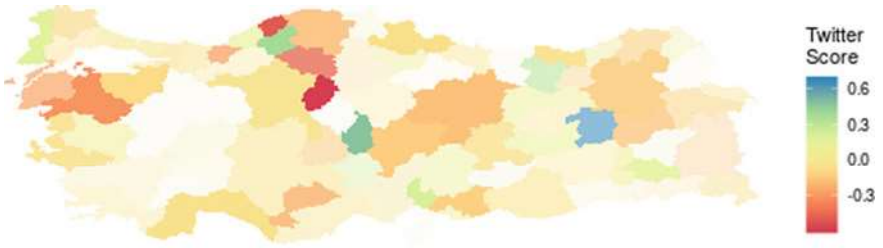

Fig. 14.5 Province-Level Average Tweet Scores from sentiment analysis. Opaqueness is adjusted for 0 value Z-score. High values indicate more positive (or less negative) sentiments. Significant differences of average twitter score estimates across provinces were not found

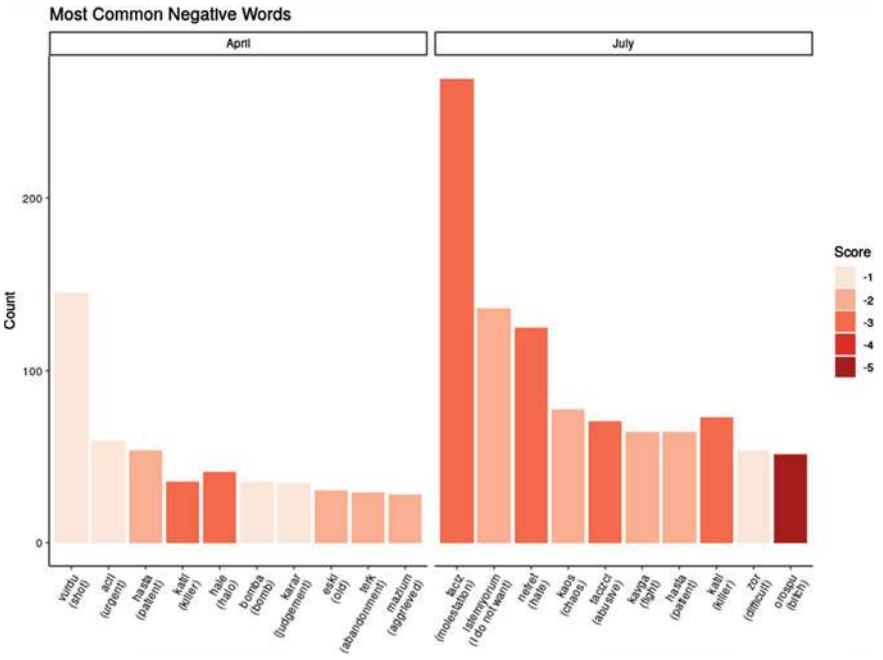

Fig. 14.6 Comparison of most common negative words in our data set of tweets about refugees for selected months

ining geographic differences in tweets (Fig. 14.5). Analysis of changes over time found that sentiment of tweets was lower in the months of June through September than in the other months (Fig. 14.3). This pattern is noteworthy in that it also appears in 2016, again with lower sentiment scores in the months between June and September. The content of the tweets was examined and the most negatively rated words for June through September drastically differed from other months, and were consistent with the way previous research found Syrians to be negatively characterized (Fig. 14.6).

Analysis of tweet sentiment, statistical model 2, at the province level using our collection of province-specific covariates was not statistically significant. While the covariates were largely in the expected direction (higher dissimilarity and urban areas led to lower predicted sentiment), our restricted sample size and noisy signal limit our ability to detect small differences in sentiment across provinces in Turkey. An increased sample size, in the form of a larger collection of tweets, would allow us to detect differences despite a noisy signal and analyze effects at a district level, where we expect measures of activity space dissimilarity to be more informative than at the provincial level.

Models for predicting calls and texts from refugees to non-refugees, statistical model 1, showed a significant positive relationship between Twitter sentiment and connectivity. As weekly Twitter sentiment scores increased, i.e., more positive text occurred in tweets about refugees, we observe higher probabilities of refugees contacting non-refugees. To evaluate the robustness of the relationship and remove potential confounding effects, we constructed a number of models with additional covariates. The effect was consistent across all models, and robust to the inclusion of other variables as seen in panel 3 of Fig. 14.7. The probability between cross-group

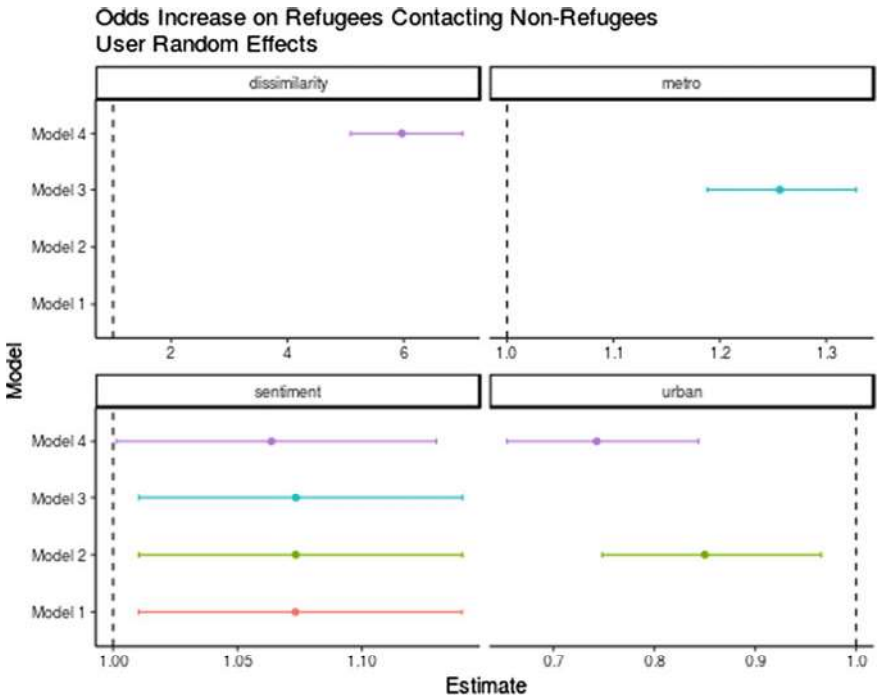

Fig. 14.7 Model Odd Ratios Coefficient Estimates for Select Covariates. Error bars not overlapping with dotted line indicate a significant result. Four models are presented in the figure on the y- axis and coefficients are placed in separate panels. Full explanation of models and covariates can be found in the Appendix

connections was larger in urban areas than non-urban, and higher when dissimilarity was higher. This pattern, however, is sensitive to the definition of urban area. The full specification of all models which follow the structure of Eq. 14.3, can be found in the Appendix along with an extended definition of each covariate and which covariates were included in each model.

14.6 Discussion

We find activity space differences between major metropolitan areas by analyzing the movements of refugees and Turkish citizens through CDR data. Meaningful differences in activity space dissimilarity exist both within and between provinces. Furthermore, the differences that we observe between locations appear to be consistent over time (Fig. 14.8). Previous research has shown that heightened segregation between groups can lead to an inability of marginalized groups to access opportunities [5] and is connected to higher rates of xenophobia, especially when related to immigrants [6]. Decreasing segregation between groups should be seen as a goal in and of itself, especially in population-dense areas where contact with other groups is more easily attainable because of spatial proximity.

In addition, we find a significant positive association between social segregation, as measured through intergroup calls from refugees to non-refugees, and sentiment of discussion of refugees on Twitter. Though the effect size that we find for the relationship is small, its presence persists across all model covariate specifications. Previous research concerning changes in the way that social media negatively discusses Syrian refugees are few [12, 15], and most often do not make connections between how changes in portrayal co-occur with increased social isolation. Our analysis finds that negative changes in sentiment toward refugees—as calculated from sentiment analysis of Twitter posts—are significantly correlated with a decrease in the probability of refugees communicating with non-refugees. Though the majority of calls and texts made by refugees go to non-refugees, it should be noted that the non-refugee group covers a broad range of individuals (Turkish citizens), groups (Turkish entities), and services (such as Arabic-speaking call centers with information on social services for refugees). Refugees rely heavily on their phones to navigate their new environment in Turkey [25]. Even small changes in the reduction of connections made by refugees to others could prove to be damaging. Events that deter refugees from connecting with non-refugees, such as changes in online portrayal and attitudes towards Syrian refugees, should be closely monitored.

The inability to detect significant differences in the average tweet score between geographic regions does not give this analysis enough signal to leverage in order to test different geographic sentiments. This does not mean that different regions tweet in a similar matter, but rather that our current resources did not allow us to capture the signal in an adequate way. There are two ways that we can potentially overcome this obstacle in future studies. One way is to use a more sophisticated process to classify tweets as positive or negative via statistical training. By labeling tweets as

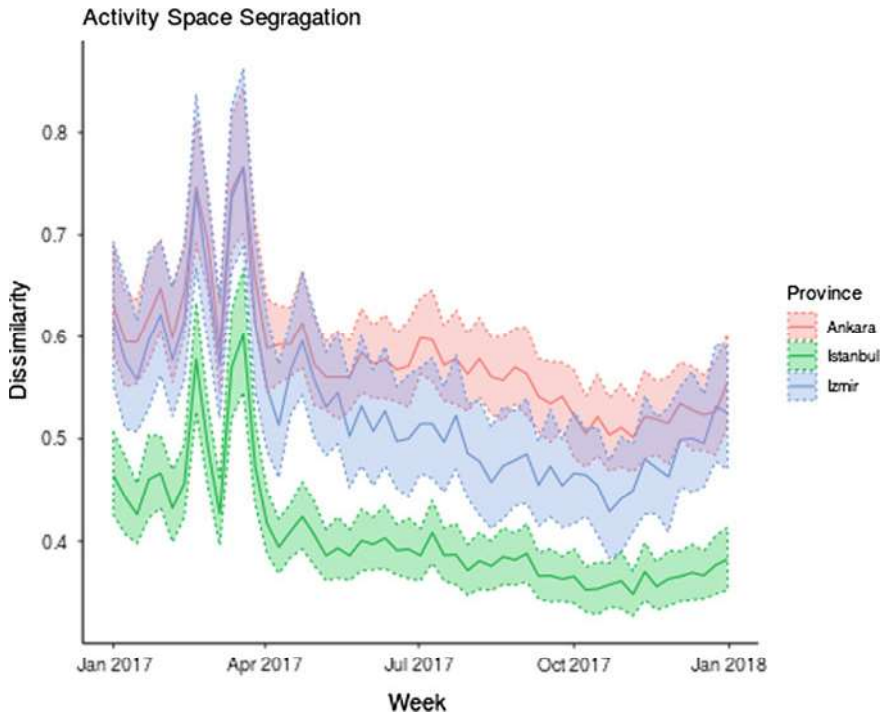

Fig. 14.8 Change in Dissimilarity by Week for Select Provinces. Uncertainty calculated from bootstrapped samples with 95% confidence intervals shown

either positive or negative via manual coding for a small set of tweets, we would be able to train a statistical model on features extracted from the text. This would allow us to focus our sentiment detection on the language that is specific to the topic of Syrian refugees. Alternatively, by increasing the sample size of our tweets, we would be able to better detect differences in signals over time and space. This could be done by using a proactive data collection strategy with the Twitter API which would allow us to collect a much greater sample than the 1% historical records provide. Another possibility for future analysis would be to remove tweets from reporting sources to filter the desired single. In the current analysis, we include all tweets from the Twitter archives that include any of a select number of words (see Appendix). By removing tweets from reporting agencies and NGOs, we may better detect public attitudes from tweets in the sentiment analysis.

14.7 Conclusion

This analysis is the first to provide comprehensive measures of segregation, both activity space and residential, between Syrian refugees and Turkish natives. We find that there are significant differences between major metropolitan areas within Turkey that are home to a significant share of the refugee population. Given that segregation has been a reported factor in the continuation of xenophobic language toward minority groups we find that it would be of interest to policymakers to continue to measure the level of both activity space and residential segregation in the near future.

Furthermore, we find that there is significant variation over time in attitudes toward refugees in Turkey on the social media platform Twitter. These variations could prove to be helpful as a gauge of changing attitudes toward Syrian refugees in light of particular events. The evidence for a relationship between segregation and changes in attitude towards Syrian refugees is limited; however, the consequences of reducing connections between Syrian refugees and Turkish natives could have dramatic consequences. Better data collection or sentiment detection could enable us to better make connections between geographic and temporal differences in sentiment and should be pursued further.

14.8 Appendix

Twitter Collection Keywords

Tweets were collected from the Twitter Archives for the period between January 1, 2017 and December 3, 2017. Any tweets that contained the following words which pertain to Syrian refugees were included in our analysis.

Suriye	mültecileri	Suriye Makedonya
Suriyeli	göç dalgasi	şişme bot göçmen
suriyeli	Suriye Yunanistan	sahil güvenlik göçmen
mülteci	Suriye Macaristan	düzensiz göçmen
mülteciler	Yunanistan'a göç	göçmen iadesi
mültecilere	Yunanistan göçmen	ÜlkemdeSuriyeliİstemiyorum

Covariate Abbreviations

Covariate	Description
sentiment	Weekly sentiment score derived from tweets about Syrian refugees in Turkey.
lrPop	Natural log population of district derived from 2014 census.
urban	The percentage of man-made land coverage from CORINE Land Coverage Database.
metroTRUE	Dummy variable where True indicates a district is in one of the top 5 urban provinces.
borderTRUE	Dummy variable indicating whether a district is in a province that borders Syria.
diss	Activity space dissimilarity at the district level calculated from a single week of data.

Model Specifications

Model 1

$$\hat{p}_{ij} = \text{logit}(\beta_0 + \beta_1 \text{ sentiment} + \zeta_i)$$

Model 2

$$\hat{p}_{ij} = \text{logit}(\beta_0 + \beta_1 \text{ sentiment} + \beta_2 \text{ lrpop} + \beta_3 \text{ urban} + \zeta_i)$$

Model 3

$$\hat{p}_{ij} = \text{logit}(\beta_0 + \beta_1 \text{ sentiment} + \beta_2 \text{ lrpop} + \beta_3 \text{ metroTRUE} + \beta_4 \text{ borderTRUE} + \zeta_i)$$

Model 4

$$\hat{p}_{ij} = \text{logit}(\beta_0 + \beta_1 \text{ sentiment} + \beta_2 \text{ lrpop} + \beta_3 \text{ urban} + \beta_4 \text{ diss} + \zeta_i)$$

Model Results Table

Model	Covariate	Estimate	Std. Error	Pr(>|z|)
Model 1	sentiment	0.07	0.03	<0.05*
Model 2	sentiment	0.07	0.03	<0.05*
Model 2	lrPop	−0.02	0.01	0.11
Model 2	urban	−0.16	0.06	<0.05*
Model 3	sentiment	0.07	0.03	<0.05*
Model 3	lrPop	−0.11	0.01	<0.05*
Model 3	metroTRUE	0.23	0.03	<0.05*
Model 3	borderTRUE	−0.01	0.02	0.64
Model 4	sentiment	0.06	0.03	<0.05*
Model 4	lrPop	0.05	0.01	<0.05*
Model 4	diss	1.79	0.08	<0.05*
Model 4	urban	−0.30	0.06	<0.05*

*$p < .05$ level

References

1. Carpenter TG (2013) Tangled web: the syrian civil war and its implications. Mediterranean Q 24(1):1–11. https://doi.org/10.1215/10474552-2018988, https://read.dukeupress.edu/mediterranean-quarterly/article/24/1/1-11/1832
2. Dinçer OB, Federici V, Ferris E, Karaca S, KiriŞci K, ÇarmIKlIEÖ, (2013) Turkey and Syrian refugees: the limits of hospitality. Tech. rep. Brookings Institute
3. İçduygu A (2015) Syrian refugees in Turkey: the long road ahead—migrationpolicy.org. Tech. rep. Migration Policy Institute, Washington. https://www.migrationpolicy.org/research/syrian-refugees-turkey-long-road-ahead
4. International Crisis Group (2016) Turkey's refugee crisis: the politics of permanence. https://d2071andvip0wj.cloudfront.net/241-turkey-s-refugee-crisis-the-politics-of-permanence_0.pdf
5. Pacione M (1999) Applied geography: principles and practice: an introduction to useful research in physical, environmental and human geography. Routledge, New York
6. Legge JS (2003) Jews, Turks, and other strangers: the roots of prejudice in modern Germany. The University of Wisconsin Press, Madison. https://uwpress.wisc.edu/books/2200.htm
7. Crush J, Ramachandran S (2009) Xenophobia, international migration and human development. Human development reports. Tech. rep. United Nations Development Programme, New York. http://hdr.undp.org/en/content/xenophobia-international-migration-and-human-development
8. Balkan B, Tok EO, Torun H, Tumen S (2018) Immigration, housing rents, and residential segregation: evidence from Syrian refugees in Turkey. Tech. rep. IZA Institute of labor Economics, Bonn
9. İçduygu A, Diker E (2017) Labor market integration of Syrian refugees in Turkey: from refugees to settlers. Göç Arasturmalari Dergisi 3(1)
10. Erdoğan MM (2018) Syrians barometer 2017. Tech. rep. İstanbul Bilgi University, Istanbul
11. Alhayek K (2014) Double marginalization: the invisibility of Syrian refugee women's perspectives in mainstream online activism and global media. Fem Media Stud 14(4):696–700. https://doi.org/10.1080/14680777.2014.935205, http://www.tandfonline.com/doi/abs/10.1080/14680777.2014.935205
12. Rettberg JW, Gajjala R (2016) Terrorists or cowards: negative portrayals of male Syrian refugees in social media. Fem Media Stud 16(1):178–181. https://doi.org/10.1080/14680777.2016.1120493, http://www.tandfonline.com/doi/full/10.1080/14680777.2016.1120493
13. Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. J Comput Sci 2(1):1–8. https://doi.org/10.1016/J.JOCS.2010.12.007, https://www.sciencedirect.com/science/article/pii/S187775031100007X
14. Flores RD (2017) Do anti-immigrant laws shape public sentiment? a study of Arizona's SB 1070 using twitter data. Am J Sociol 123(2):333–384. https://doi.org/10.1086/692983, https://www.journals.uchicago.edu/doi/10.1086/692983
15. International Crisis Group (2018) Turkeys Syrian refugees: defusing metropolitan tensions. https://www.crisisgroup.org/europe-central-asia/western-europemediterranean/turkey/248-turkeys-syrian-refugees-defusing-metropolitan-tensions
16. We Are Social (2018) Digital Report 2018. https://digitalreport.wearesocial.com/
17. Salah A, Pentland A, Lepri B, Letouzé E, Vinck P, de Montjoye Y, Dong X, Dağdelen Ö (2018) Data for refugees: the D4R challenge on mobility of Syrian refugees in Turkey. arXiv preprint arXiv:180700523
18. Erdoğan MM, Şener B, Sipahioğlu E, Kavukçuer Y, Basçeri EY (2017) Urban refugees from detachment to harmonization. Tech. rep. Marmara Municipalities Union's Center for Urban Policies. https://mmuraterdogan.files.wordpress.com/2016/06/mmu-urban-refugees-report-2017_en.pdf
19. Land Cover Change (LCC) 2006–2012, Version 18.5 (2016) https://land.copernicus.eu/pan-european/corine-land-cover/lcc-2006-2012?tab=metadata

20. Center for International Earth Science Information Network (CIESIN)—Columbia University (2016) Gridded population of the world, version 4 (GPWv4): population density. Tech. rep. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://beta.sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density
21. Lee DT, Schachter BJ (1980) Two algorithms for constructing a Delaunay triangulation. Int J Comput Inf Sci 9(3):219–242. https://doi.org/10.1007/BF00977785, http://link.springer.com/10.1007/BF00977785
22. Wong DW, Shaw SL (2011) Measuring segregation: an activity space approach. J Geogr Syst 13(2):127–145. https://doi.org/10.1007/s10109-010-0112-x, http://www.ncbi.nlm.nih.gov/pubmed/21643546
23. Nielsen FÅ (2011) AFINN. Informatics and mathematical modelling. Tech. rep. Technical University of Denmark, Kongens Lyngby. http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010
24. Toomet O, Silm S, Saluveer E, Ahas R, Tammaru T (2015) Where do ethno-linguistic groups meet? How copresence during free-time is related to copresence at home and at work. PLOS ONE 10(5):e0126093. http://dx.plos.org/10.1371/journal.pone.0126093
25. Narli N (2018) Life, connectivity and integration of Syrian refugees in Turkey: surviving through a smart phone. Quest de Commun 33(1):269–286. https://www.cairn.info/revue-questions-de-communication-2018-1-p-269.htm