

Social Media based Analysis of Refugees in Turkey

Abdullah Bulbul, Cagri Kaplan, and Salah Haj Ismail

Ankara Yildirim Beyazit University, Türkiye,
abulbul@ybu.edu.tr
<http://ybu.edu.tr/abulbul>

Abstract. In this paper, we propose a method to find out public social media accounts of refugees, and trace them back to infer temporal and spatial events. These data will have many future applications in planning and design of solutions for their problems and needs, starting from legislations and social ones, to architectural and housing needs. It will lead to better understanding of the obstacles they face which they fear to express in direct interviews and inquests. In this first application, we present our method to retrieve information from social media, share the characteristics of the data gathered and perform initial analysis with a discussion of future opportunities and difficulties.

Keywords: information retrieval, social media, refugees

1 Introduction

After Arab spring and following wars, Turkey accommodates more than 3 millions of Syrian refugees [2]. Being also one of the accelerators of the Arab spring, social media usage have accompanied majority of the events [8]. Having a dataset of refugees in Turkey has the potential to facilitate a number of future studies.

To understand the obstacles and difficulties the refugees face in their refuge countries, it is necessary to conduct questionnaires and interviews, which is difficult and even forbidden in some countries. Thus, an indicative database could be established through data retrieval from social networks, which can serve for better understanding of refugees needs, and the future planning to bridge the gap between what they have and need for better integration in their new societies.

This paper presents a tool and methodology to collect data from Twitter accounts, refine them to define the Syrian refugees accounts, then trace their tweets back to collect a database to be analyzed by time and location across Turkey, defining the trend topics occupied the refugees minds to react with, tweeting and expressing their opinions both on the issues of their home country and new refuge country. In results, those trends were shown in a graphical presentation which enables better understanding of their needs with better planning for future intervention to solve social, economical and maybe political refugee issues and aspects in Turkey the country hosting highest number of refugees in the world today[1].

As a result of the conflict occurred in Syria, many citizens have migrated abroad since 2010. As an example of refugees related work, the study in [5] explores the ideas of users towards refugees by means of collected tweets including #refugeesnotwelcome hashtag. The work in [7] also exploits tweets including same hashtag to understand the portrayal of male Syrian refugees on social media. To understand outcomes of an immigrant related event on society, the work in [6] studies the prediction of attitudes of US Twitter users towards Islam and Muslims subsequent to the tragic Paris terrorist attacks that occurred on November 13, 2015. As an interesting research, the study in [3] shows the propagation power of Twitter on users via "seminar users" who are social media users engaged in propaganda in support of a political entity.

2 Used Method

To retrieve information from refugee related social media accounts, we analyzed public Twitter activity using Twitter API¹. The process is divided into four steps: Firstly, we tried to figure out a method to define the accounts of Syrian refugees in Turkey. Secondly, we traced back those chosen users' accounts. Thirdly, we analyzed the data and classified it into groups by location and year. Finally, a trend analysis was conducted to reveal the most important issues the refugees discussed for each year.

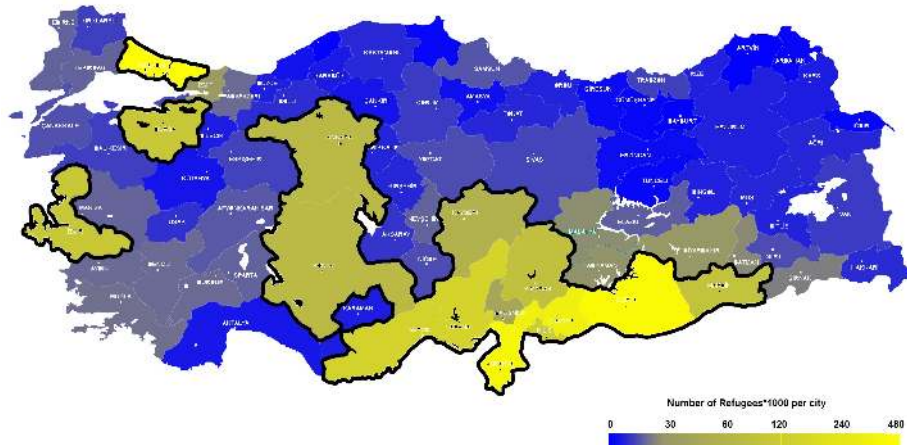


Fig. 1: Distribution of Syrian refugees in Turkish cities generated according to the statistics in [4].

Determining the accounts. The first heuristic we can use to determine the refugees is to check their languages. In Twitter, users are associated with a language and the majority of refugees (95% [1]) have Arabic as their mother tongue and refugees are forming 90% of the registered foreigners in Turkey, so we assume that Twitter accounts in Turkey who uses Arabic as the main language are probably refugees'.

¹ <https://developer.twitter.com/en/docs>

Syrians are the majority of refugees in Turkey, therefore, we checked the number of registered Syrian refugees in each city. Figure 1 shows the distribution of Syrian refugees in Turkey. Using a logarithmic color code, the image is generated according to the statistics in [4]. From Figure 1, it is clear that the majority of the refugees are hosted in specific cities. The bold border shows the regions chosen in our study. The cities in these regions accommodates 90% of Syrian refugees while they only have 48% of the total Turkish population. Table 1 shows the number of refugees in chosen cities and the ratio of refugees to the city populations.

Table 1: Main Cities Hosting Refugees

	Adana	Ankara	Bursa	G.Antep	Hatay	Istanbul	Izmir	K.Maras
N.Refugees	150790	73198	106000	329670	384024	479555	108306	90100
Ratio	6.85%	1.37%	3.68%	16.70%	24.69%	3.24%	2.58%	8.11%
	Kayseri	Kilis	Konya	Mardin	Mersin	Osmaniye	S.Urfa	Total
N.Refugees	59938	124000	73445	94340	146931	43773	420532	2685669
Ratio	4.34%	95.15%	3.40%	11.85%	8.28%	8.38%	21.67%	6.28%

Twitter API lets searching for recent (seven days) tweets according to keywords, location, and used language. To avoid a biased collection of data, we don't provide any keywords. We search for Arabic tweets in specific locations wherein refugees are accommodated intensely.

For practical purposes we limited the search to 1000 tweets per region. Then, we extracted the individual user IDs posting these tweets. As one account can post multiple tweets, we have less amount of user accounts than the number of tweets. We performed this procedure twice with 4 days interval to increase the number of users. As a result we collected a total of 5707 twitter users who were active recently. Table 2 shows the number of discovered users in each region.

Table 2: Data collected: users- Tweets per city

	Adana	Ankara	Bursa	G.Antep	Hatay	Istanbul	Izmir	K.Maras
N.Users	160	1050	2023	535	12	1183	78	35
N.Tweets	127031	2297028	4543793	1139878	13874	2910948	107263	13538
	Kayseri	Kilis	Konya	Mardin	Mersin	Osmaniye	S.Urfa	Total
N.Users	62	3	137	30	269	22	108	5707
N.Tweets	55275	558	158845	24130	461598	11811	156736	12022306

Tracing back and filtering out irrelevant accounts. Twitter API lets accessing up to last 3200 of a user's tweets including retweets. Therefore, we gathered tweets of each user we extracted in the previous step until the limit is reached or there is no more tweet from that user. In one query, it is possible to access only 200 tweets, thus, we run multiple queries to collect the maximum possible number of tweets. Table 2 includes the number of tweets collected from each city.

Among these accounts, there are ones which do not belong to individual users but to press or companies for instance. These accounts mostly post with a very high frequency including a big ratio of retweets. This information is helpful to differentiate the users we are interested in to analyze in this study. To analyze the development of refugee related issues in post-”Arab Spring” years, we focus more on the users for whom we have data that covers those years. Table 3 shows the number of accounts with the date of their oldest tweets collected. We have selected the users that were active in or before 2014.

Table 3: Number of users with their oldest known activity

Year	2018	2017	2016	2015	2014	2013	2012	2011	2010
N.Users	876	2997	815	397	207	194	147	63	22

Analysis. Firstly we have excluded the tweets from 2018, only tweets up to the end of 2017 were analyzed, since in 2018, millions of tweets (3,678,739) were retrieved in less than 20 days which are unreliable, due to the fact that tweeting 3200 tweets in those days means more than 160 tweets per day, implying that these tweets are either not coming from a real user, or not expressing an individual user opinion. Moreover, we have excluded the Twitter accounts created or have all the activities after 2014 since these accounts are incapable to reflect the continuous change of refugees’ conditions and needs.

With the aforementioned approach, out of 5707 accounts and 12 million tweets, we have refined the accounts into 633 and the tweets to almost 800 thousands, and classified these tweets into: tweets (435,378) and RT(336,753) using the tweets only we have created the word clouds seen in Figure 2. The analysis shown is an example which was conducted for the total tweets in Turkey.

3 Preliminary Results and Discussion

The word clouds shown in Figure 2 are created to show the trends and most frequent words in the collected data set. We used an Arabic light stemmer [9] to select the correct words and avoid pronouns, conjunction, suffixes and non useful or meaningful words etc. A general view over the analyzed five years shows that the main topic that occupied the refugees minds was the world’s reaction to their tragedy, words like: world, hearts and friend express the disappointed *hearts* of Syrians towards the *world* which called himself Syrian peoples *friend*. Moreover, with all the suffer of the refugees life during the war, *peace* was the third most repeated word in their tweets, which represents clearly their eagerness to end the war and establish permanent peace. While with a yearly based analysis, we noticed the change in topics across the years from deep involvement in Syrian news and issues in the first four years to appear issues related with obstacles in Turkey such as *work*, *study*, dreaming of *returning*, *beauty* and achievements of *Turkey* and of course the coupe attempt and taking decision of immigrating to Europe or staying in Turkey...etc. Acts of same accounts such as using Turkish words in Arabic alphabet like *Para* which means (Money), then using Turkish



2012-2017 (En)



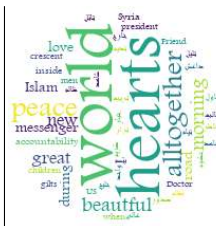
2012-2017 (Ar)



2012 (En)



2012 (Ar)



2015 (En)



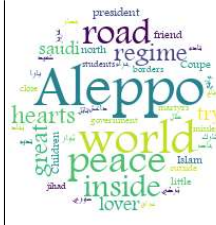
2015 (Ar)



2013 (En)



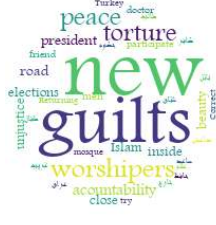
2013 (Ar)



2016 (En)



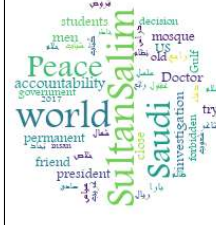
2016 (Ar)



2014 (En)



2014 (Ar)



2017 (En)



2017 (Ar)

Fig. 2: Generated wordclouds for years 2012 to 2017. Same wordclouds for each year presented: Original Arabic words (right), 20 most frequent words translated to English (left).

language and alphabet in the tweets, reflects that those users have started to learn Turkish, as their integration process in the society has become deeper and stronger since their stay took longer than they expected at the beginning. In a location based analysis even clearer differences in refugees situation and trending issues between Turkish cities could be concluded. Different fields and topics analysis can be applied to this data set in order to retrieve more information about specific issues, such as services to refugees, legislative and logistics of their stay in the refuge countries, and reactions to politics change and treatment in hosting countries. The preliminary analysis shows accounts that tweets according to events, while other accounts are created to form the public opinions about specific political topics. This provides wide potentials for deeper future analysis to better understanding the conditions of refugees, and to enhance the public plans for their integration. To facilitate further analysis, we share the dataset and our analysis online ². To understand the cultural and social value of cultural heritage issues, future work will focus on analysis of cultural elements which the refugees were aware of and concerned about. To represent the collective identity of a community or nation, that has been forced to leave and refuge in other countries, by visual means is another future work direction. The interaction of locals with refugees' issues will also be studied and analyzed to compare the different treatment of the two communities in the hosting country.

References

1. United nations refugee agency. <http://www.unhcr.org/figures-at-a-glance.html> (2017), accessed: 2018-01-19
2. 3rp:syria regional refugee response inter-agency information sharing portal. <http://data.unhcr.org/syrianrefugees/country.php?id=224> (2018), accessed: 2018-01-19
3. Darwish, K., Alexandrov, D., Nakov, P., Mejova, Y.: Seminar users in the arabic twitter sphere. In: International Conference on Social Informatics. pp. 91–108. Springer (2017)
4. Esnek, F.: Basin ilan kurumu, hangi ilde ne kadar suriyeli var? iste il il liste. <http://www.bik.gov.tr/hangi-ilde-ne-kadar-suriyeli-var-iste-il-il-liste/> (2017), accessed: 2018-01-19
5. Kreis, R.: # refugeesnotwelcome: Anti-refugee discourse on twitter. *Discourse & Communication* 11(5), 498–514 (2017)
6. Magdy, W., Darwish, K., Abokhodair, N., Rahimi, A., Baldwin, T.: #isisisnotislam or #deportallmuslims?: Predicting unspoken views. In: Proceedings of the 8th ACM Conference on Web Science. pp. 95–106. WebSci '16, ACM, New York, NY, USA (2016), <http://doi.acm.org/10.1145/2908131.2908150>
7. Rettberg, J.W., Gajjala, R.: Terrorists or cowards: negative portrayals of male syrian refugees in social media. *Feminist Media Studies* 16(1), 178–181 (2016)
8. Stepanova, E.: The role of information communication technologies in the arab spring. PONARS Eurasia Policy Memo 159, 12–16 (2011)
9. Zerrouki, T.: Tashaphyne, arabic light stemmer (2012), <https://pypi.python.org/pypi/Tashaphyne/0.2>

² http://ybu.edu.tr/abulbul/custom_page-283-refugees