Real-Time Monitoring of the Venezuelan Exodus through Facebook's Advertising Platform

J. Palotti[1], N. Adler[2], A. J. Morales[3], J. Villaveces[4], V. Sekara[2], M. Garcia Herranz[2], M. Al-Asad[5], I. Weber[1*]

[1]Qatar Computing Research Institute, HBKU
E-mails: jpalotti,iweber@hbku.edu.qa
[2]UNICEF
E-mail: nadler,vsekara,mgarciaherranz@unicef.org
[3]MIT Media Lab
E-mail: alfredom@mit.edu
[4]iMMAP Colombia
E-mail: jvillaveces@immap.org
[5]Global Protection Cluster, Geneva, Switzerland.
E-mail: alasad@unhcr.org
*To whom correspondence should be addressed; E-mail: iweber@hbku.edu.qa.

March 29, 2019

Abstract

Venezuela is going through the worst economical, political and social crisis in its modern history. Basic products like food or medicine are scarce and hyperinflation is combined with economic depression. This situation is creating an unprecedented refugee and migrant crisis in the region. Governments and international agencies have not been able to consistently leverage reliable information using traditional methods. Therefore, to organize and deploy any kind of humanitarian response, it is crucial to evaluate new methodologies to measure the number and location of Venezuelan refugees and migrants across Latin America. In this paper, we propose to use Facebook's advertising platform as an additional data source for monitoring the ongoing crisis. We estimate and validate national and sub-national numbers of refugees and migrants and break-down their socio-economic profiles to further understand the complexity of the phenomenon. Although limitations exist, we believe that the presented methodology can be of value for real-time assessment of refugee and migrant crises world-wide.

Technical Report

The current economic and political crisis in Venezuela has led to an outpouring of refugees and migrants from the country. As of February 2019, the Regional Inter-Agency Coordination Platform for Refugees and Migrants from Venezuela (R4V) estimates that there are 3.4 million people who have left Venezuela and who currently live in countries in Latin America and the Caribbean region [1]. According to the United Nations High Commissioner for Refugees (UNHCR), "this is the largest exodus in the recent history of Latin America"[1]. This has prompted humanitarian interventions from governments, UN agencies and civil society. Targeted efforts, however, have been hampered by missing, outdated or incorrect data concerning the (i) absolute number, (ii) spatial distribution, and (iii) socio-economic composition of Venezuelan refugees and migrants.

Attempts to measure the size of the population displacement have been conducted by governments in the receiving countries. Colombia, for example, ran the *Registro Administrativo de Migrantes Venezolanos* (RAMV) from April to June, 2018[2], a registry campaign in which Venezuelans living in Colombia could voluntarily identify themselves at specific registration points designated by the government. While this approach provides a snapshot of the crisis, it is limited in many ways. First, it only provides a single snapshot of refugee and migrant stocks and is therefore inadequate to capture sudden flows, e.g., due to an unforeseen deterioration of the situation in Venezuela. Second, the effort of staffing one thousand registration posts for 61 days comes with a significant economic burden. Third, and perhaps most crucially, it relies on self-reported data, which is likely to be incomplete, leading to a considerable under-count of the affected population. One reason for the lack of participation in voluntary registration campaigns such as RAMV is the perceived lack of benefits to registering, in particular for refugees and migrants with an irregular status.

For these reasons, which largely apply to displacement and migration monitoring in general, researchers have explored non-traditional data sources for monitoring trans-national mobility, including IP addresses of email users [2], geo-tagged tweets [3], satellite data to count structures

[1] `https://www.unhcr.org/venezuela-emergency.html`, retrieved on Feb 26, 2019.

[2] Available in Spanish at `https://data2.unhcr.org/en/documents/download/64101`, last accessed on Feb 26, 2019.

in refugee settlements [4] and Facebook's advertising audience estimates [5]. In this work, we build on [5] and show how data from Facebook's advertising platform can be used to supplement the monitoring of the Venezuelan refugee and migration crisis by providing insights into (i) the temporal trends of refugee and migrant flows, (ii) the spatial distribution of Venezuelan nationals in their host countries, and (iii) the socio-economic makeup of these communities.

Facebook's advertising platform was designed to support targeted advertisements based on a large number of user attributes, including both self-reported and inferred attributes (See Fig-S1). As an example, an advertiser willing to launch an advertisement campaign on Facebook can choose to selectively target their advertisement to people who, according to Facebook's classification, (i) are aged 13 and above, (ii) are currently living in the Colombian department of Norte de Santander, (iii) previously used to live in Venezuela, and (iv) use an iOS device to access Facebook. Before the advertisement campaign is launched, Facebook provides the advertiser with an estimate of the number of monthly active users (MAUs) matching the provided targeting criteria. In this concrete example, the estimate is 3,000 MAUs (as of February 24, 2019). Additional targeting criteria include education level, relationship status, and topical interests. These estimates of user counts are available free of charge through the Facebook Graph API[3].

To understand if, despite its limitation to Facebook users and not the general population, these estimates of user counts are capturing the magnitude of the actual migration, Fig. 1 shows a comparison for the general "Facebook users aged 13 and above who used to live in Venezuela" with the most recent official estimates. Figs. 1A and 1B show the estimates of Venezuelans living in Latin American countries according to the latest report from the R4V [6] (left) and Facebook Advertisement (right), both from the same period, January 15, 2019[4].

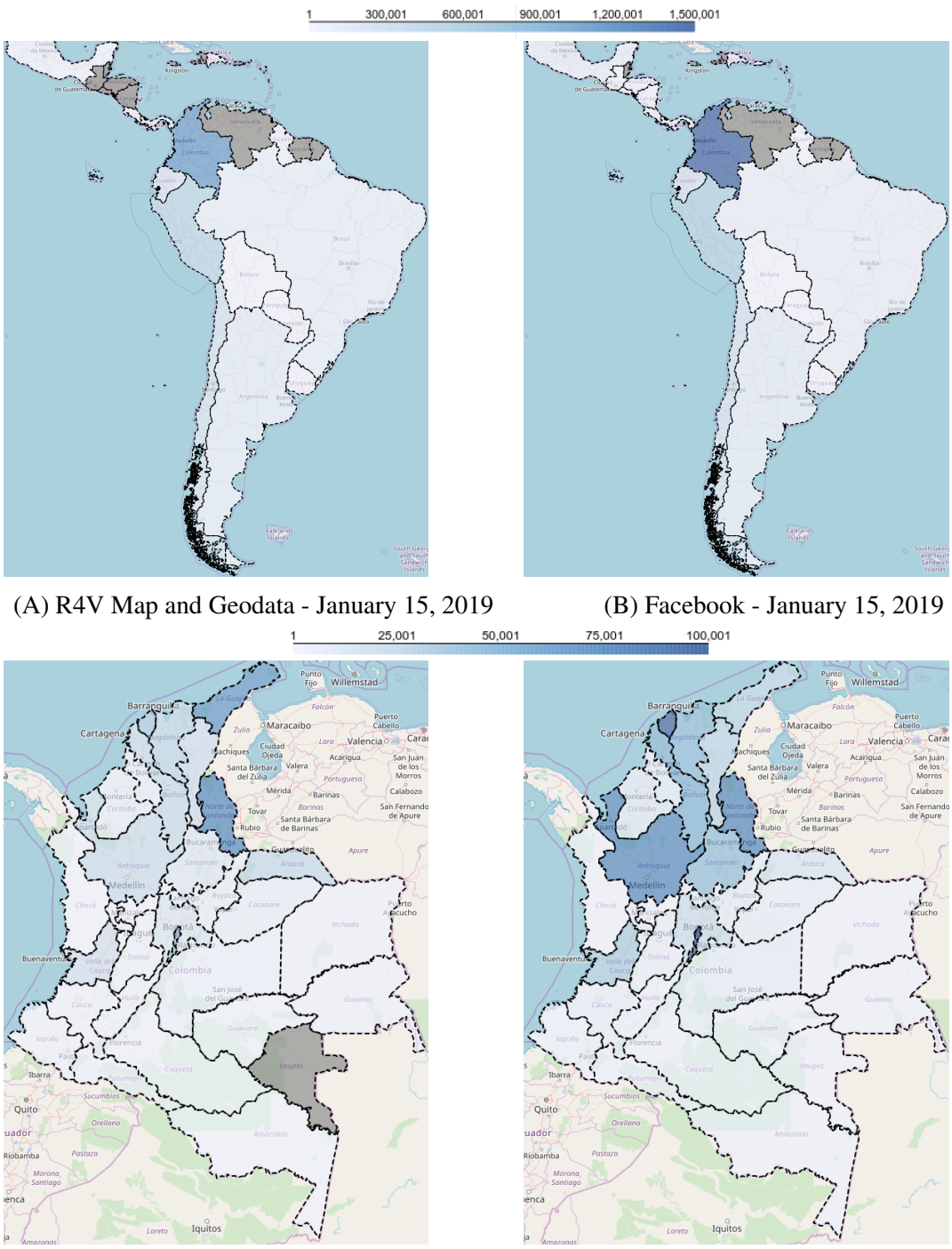
The two data sources share a similar spatial distribution with a Pearson and Kendall correlation across the different countries of, respectively, $r = 0.99$ and $\tau = 0.79$ (n = 17, $p < 0.0001$ for both metrics). In January 2019, R4V Map and Geodata reported a total of 2.7M Venezuelan nationals across 17 countries [6], while there were an estimated 3.2M Facebook users who previously lived in Venezuela now living in of these countries.

At the sub-national level, Figs. 1C and 1D compare the estimates of Venezuelan nationals (left, according to RAMV) and Facebook users who used to live in Venezuela (right, according to Facebook). The Pearson correlation of $r = .57$ (n = 31, $p = .0008$) between RAMV and Facebook is smaller than for the continental data above. The decrease in the Pearson correlation coefficient can be noted in Fig. 1D, in which Facebook estimates more refugees and migrants from Venezuela than RAMV for several departments away from the Venezuela-Colombia border. However, a Kendall's rank correlation of $\tau = .71$ (n = 31, $p < .0001$) shows that the agreement for the relative rank among the departments of both distributions is high. Thus, even if the estimates from RAMV were perfect, Facebook-derived estimates would still be a useful tool to discover where the density of migrants is higher, allowing a better targeted management of the refugee and migration crisis.

Concerning the temporal evolution of the Venezuelan refugee and migration crisis, Facebook's advertising platform is more limited as it does not offer any historical information. For example, one cannot obtain estimates on the number of Facebook users who previously lived in Venezuela

[3] `https://developers.facebook.com/docs/marketing-apis/`

[4] Countries are shown in grey when no estimates are available (Fig. 1A) or when Facebook's returned estimate is smaller than the minimum resolution of 1,000 monthly active users (Fig. 1B).



(A) R4V Map and Geodata - January 15, 2019

(B) Facebook - January 15, 2019

(C) RAMV - April 6 to June 8, 2018

(D) Facebook - June 8, 2018

Figure 1: Estimates of Venezuelan refugees and migrants in South America (Figs. 1A and 1B) and Colombia (Figs. 1C and 1D) from different data sources.

and who lived in Cúcuta, Colombia, in August 2014. Temporal trends can, however, be traced through repeated data collections, eventually building up a historic repository. Fig. 2 shows Facebook-derived estimates for eight countries in Latin America compared to estimates by the

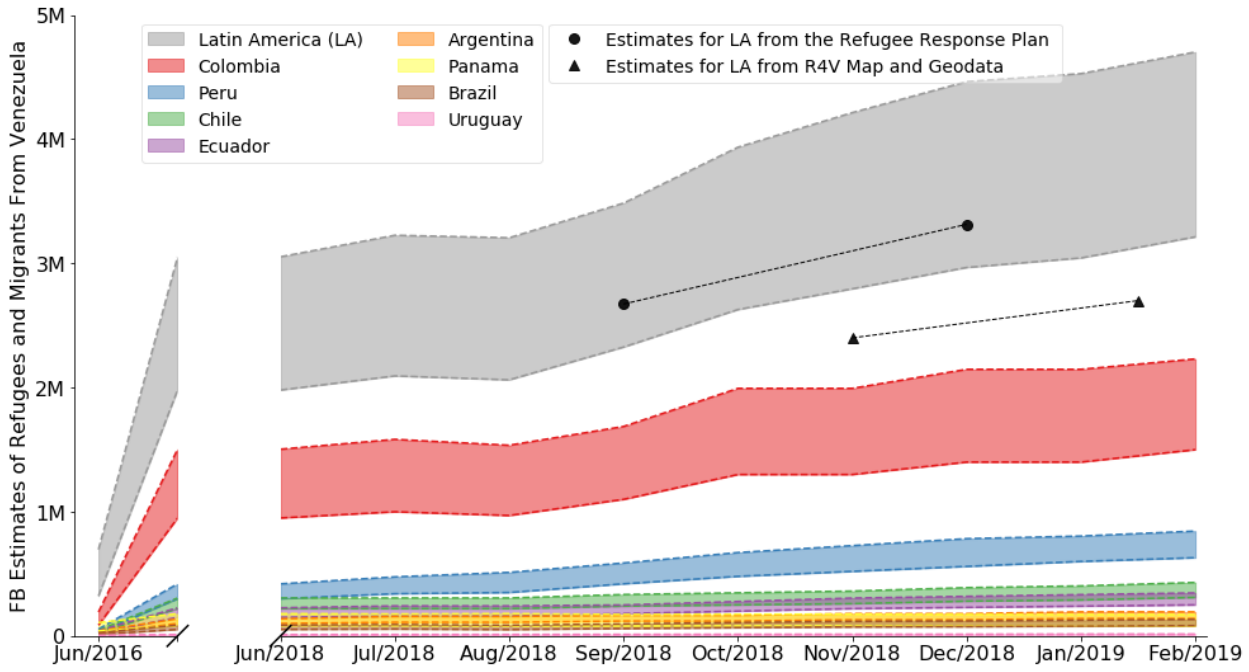


Figure 2: Migration trends for Venezuelans in Latin America. Lower estimates are the raw estimate from Facebook Marketing API, whereas upper estimates take into consideration a correction factor for Facebook penetration in the host countries. Estimates for refugees and migrants from Venezuelans in Latin America compiled in the R4V Map and Geodata [8, 6] and the Refugee Response Plan [7] are shown for comparison.

Regional Inter-Agency Coordination Platform for Refugees and Migrants from Venezuela (R4V). Note that even though both the "Response Plan" [7] and the "Map and Geodata" [8, 6] are released by R4V, their numbers are not fully consistent. The lower bounds bounds for the bands in Fig. 2 are the raw population estimates from Facebook, i.e., estimates for the number of monthly active Facebook users aged 13 and above who used to live in Venezuela but who now live in one of the host countries. The upper bounds represent estimates corrected for Facebook penetration, assuming that Venezuelan refugees and migrants are as likely to join Facebook as the population in the host country.[5] The estimates from June 2016 were originally collected for [5].

Based on the spatial (Fig. 1) and temporal (Fig. 2) comparison to the best available data from R4V Map and Geodata, the Response Plan and RAMV, the estimates for the number of Facebook users who used to live in Venezuela is a useful proxy signal for the number of Venezuelan refugees and migrants. Note that the lack of reliable "gold standard" data to use for validation is the main motivation to consider non-traditional data sources to triangulate existing ones.

Whereas the focus above was on validating, where possible, estimates obtained from Facebook, the following analyses focus on obtaining estimates for aspects where no comparable data exists. These include (i) sub-national estimates for the spatial distribution and (ii) insights into the socio-economic status of Venezuelan refugees and migrants in different host countries.

Figure 3 shows estimates, obtained from Facebook, for the spatial distribution of Venezue-

[5]See supplementary materials for details on how the Facebook penetration in the host countries is taken into account.

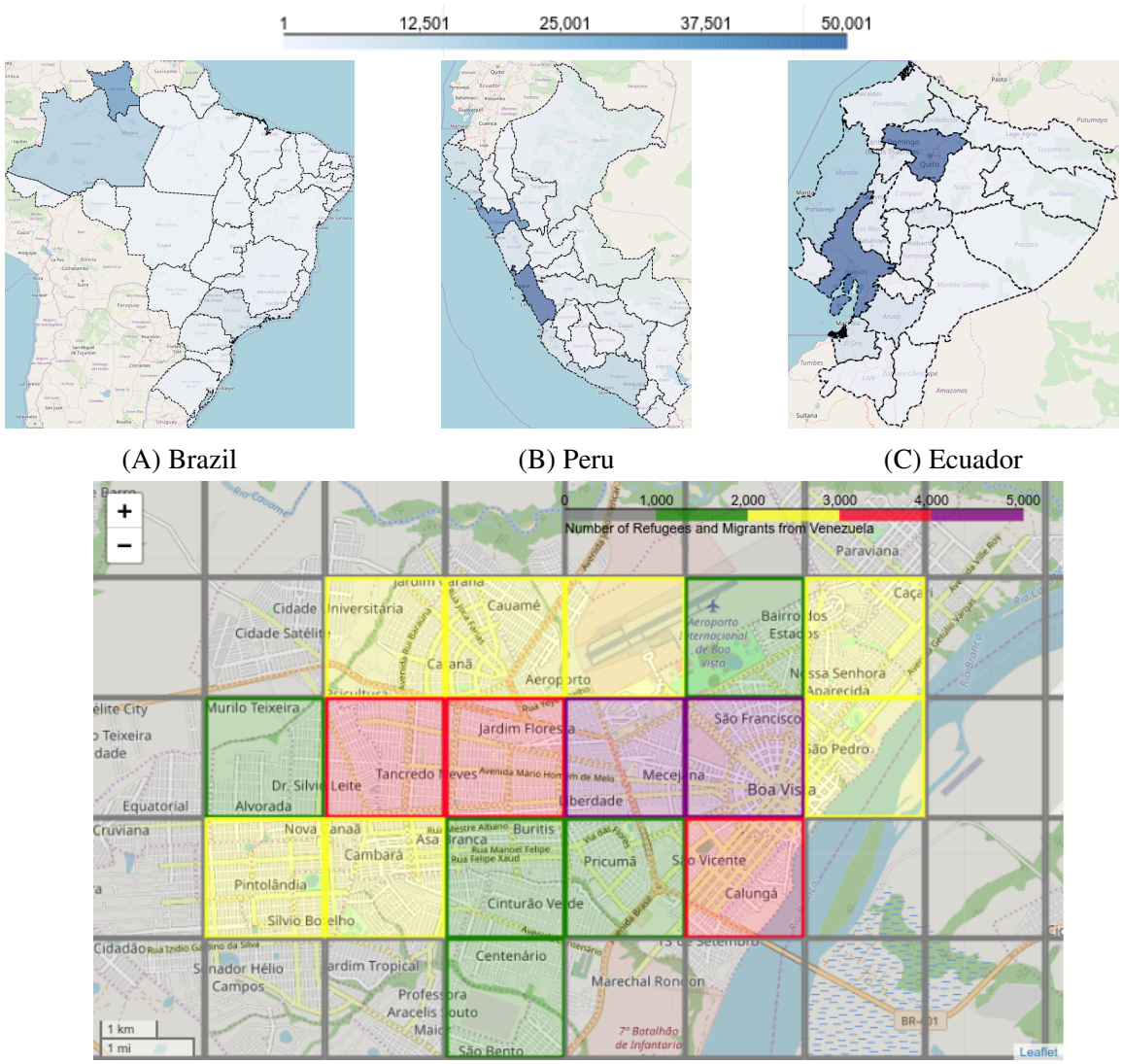
lan refugees and migrants at the highest sub-national administrative level, i.e., across provinces ("provincias" in Peru and Ecuador) or states ("estados" in Brazil). For example, Fig. 3A shows that, based on Facebook-derived estimates, around 75% of the refugees migrants from Venezuela in Brazil are in two states on the Brazil-Venezuela border, Roraima and Amazonas. Further in the South of Brazil, the richer states of Sao Paulo and Rio de Janeiro are home to 12% and 5% respectively of the Facebook users who previously lived in Venezuela. Likewise, Figs. 3B and 3C provide a state-level analysis of the estimated spatial distribution of Venezuelan refugees and migrants across Peru and Ecuador. Finally, to illustrate the spatial resolution that can be obtained via Facebook's marketing API, Fig. 3D shows a breakdown of the Brazilian city of Boa Vista in the state of Roraima, where most of the migrants from Venezuela are currently located. Figure 4 shows Facebook estimates for refugees and migrants from Venezuela in all American countries in which the estimates are bigger than the minimum resolution of 1,000 monthly active user.

Estimating the absolute number and the spatial distribution of Venezuelan refugees and migrants are a top priority to quantify the magnitude the crisis and to plan an appropriate humanitarian response. However, insights into their socio-economic status and how it compares to the host population are also important, in particular to anticipate potentially hostile sentiments from the host population. To illustrate how Facebook's audience estimates can be used for this purpose, Fig. 5 provides a socio-economic analysis of the Venezuelan population in Latin American countries, analyzing their self-reported education level (on top) and their inferred income (below). Figs. 5A and 5B report, respectively, on the self-declared education level and estimated average income per capita of refugees and migrants from Venezuela. While this information is hard to validate due to the lack of official data, there are sources that the results can be contrasted with. For example, historically countries like Panama and Costa Rica have received wealthier and educated Venezuelans [9], while countries like Colombia or Peru are recently receiving poorer, less educated ones [10, 11]. In Chile, the number of professionals and educated refugees and migrants that seem to be underemployed is consistent with recent UN Venezuelan migrant reports [12].

In this work, we showed the benefits of triangulating emerging sources of data, such as Facebook's advertising data, to supplement official refugee and migration data. In the context of the Venezuelan crisis, Facebook's advertising data has proved valuable because of its low latency (days not months), low acquisition cost (only programming efforts), high spatial resolution (sub-city resolution), and possibility to disaggregate by socio-economic status (education level and inferred income), as well as the remote sensing capabilities (no need to enter unstable regions). The global reach of Facebook operations also provides an advantage over the usage of aggregate mobile phone data such as call detail records (CDR). While CDR have been used to successfully map population distribution and mobility [13, 14, 15], they are typically limited to a single country as aggregating CDR data across countries comes with both technical and legal difficulties. This limits their use to study *cross-border* displacements [16].

A concrete example of the operational benefits of our methodology is the understanding of the spatial distribution of Venezuelan refugees in Brazil (see Fig. 3A). Knowing this distribution helped redefine and amplify the geographical scope of humanitarian and longer-term development interventions beyond the border-crossing area. The recognition of the national scale of the crisis was particularly relevant for UNICEF to develop anti-xenophobia campaigns using Facebook's chatbots.

Despite the advantages outlined above, it is important to acknowledge limitations of using non-traditional data sources [17]. One key limitation is the dependence on Facebook's inaccessible



(A) Brazil (B) Peru (C) Ecuador

(D) Boa Vista, Roraima, Brazil

Figure 3: Raw estimates of Venezuelan refugees and migrants in regions where no other data is available. Data collections done through Facebook Marketing API on February 18, 2019.

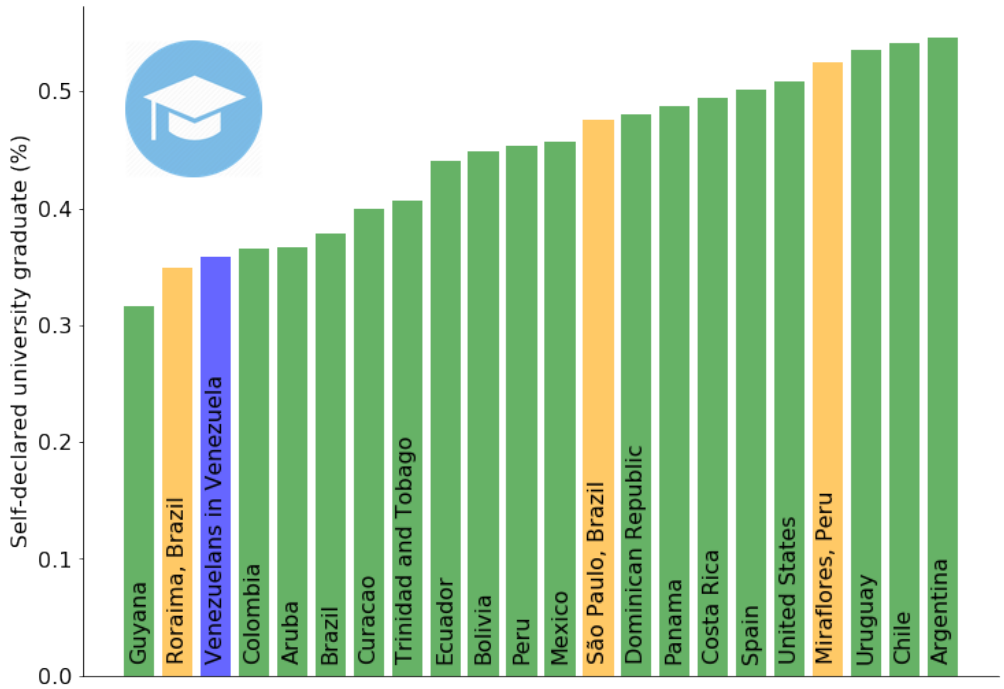
algorithm for identifying users' previous countries of residence. Although hints of this procedure can be obtained from academic work published by Facebook researchers [18], indicating that both the self-declared "home" location and social network structure play a role, the exact set of features or the evaluation used for the inference remains hidden. Regarding privacy concerns, the risks are relatively small as only anonymous and aggregate data is obtained–namely the number but not the identities of Facebook users matching provided targeting criteria. Previous researchers who identified privacy leaks in Facebook's advertising platform [19] had used so-called "custom audiences", built around mobile phone numbers or email addresses of known users, which is not done for this line of work.

Another technical limitation is the lack of availability of historic data through the advertising platform, meaning that temporal trends can only be inferred from building up a repository over

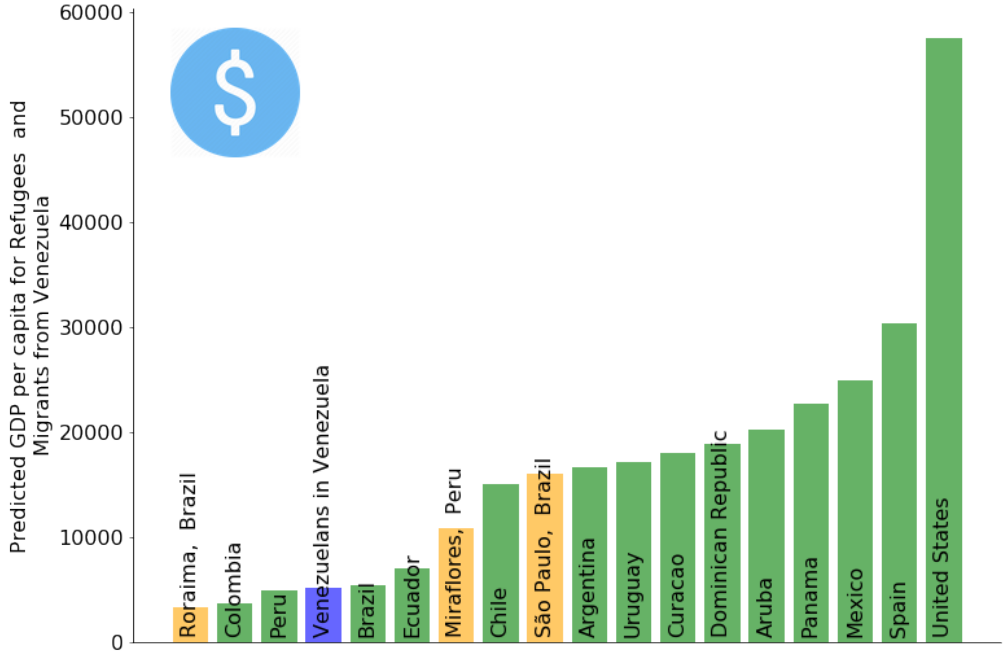
Figure 4: Number of Facebook users likely to be Venezuelan refugees and migrants, living in different host countries in February 18th, 2019. Refugees and migrants not on Facebook cannot be directly captured by our method and true numbers are likely to be higher. Estimates for the relative spatial distribution of refugees and migrants are more robust than the absolute estimates.

time. Monitoring long-term temporal trends of any platform gets further complicated by typical changes of market shares and usage patterns over time, requiring re-validating and re-calibrating models built on top of them.

Lastly, there is a risk when using digital traces to monitor humanitarian crises to exclude affected people without access to digital technology – potentially the most disadvantaged – who will not leave traces and hence remain uncounted. In our validation we do however observe that de-



(A) Percentage of university graduate refugees and migrants from Venezuela in different countries.



(B) Estimated nominal GDP per capita for refugees and migrants from Venezuela. The linear model created to generate these estimates is based on the percentage of iOS devices for the host population in each country. Details in the supplementary materials.

Figure 5: Socio-economic data for Venezuelan refugees and migrants. Data collected through Facebook Marketing API on February 18, 2019.

spite the selection bias and potential noise of the algorithm to infer users' previous countries of residence, the estimates derived from Facebook are close to the best official estimates. This is in line with the observation by Wesolowski et al. that despite the substantial bias in phone ownership, mobility patterns derived from mobile phone data are surprisingly robust [20].

Despite the general good agreement, looking at the differences between estimates derived from Facebook and official estimates reveals important patterns of bias. In Fig. 6A we illustrate how the GDP per capita of regions relates to over- and under-estimation bias. We specifically focus on Colombia because of the recent RAMV survey, which currently is the best existing ground truth estimate. Whereas Facebook in general overestimates the number of migrants and refugees, it *underestimates* for certain of the less wealthy departments, such as La Guajira and Vichada. This might, in part, be caused by the geographic proximity of these departments to the Venezuelan border (see Fig. 6B), where RAMV efforts might have been more exhaustive. But there is also good reason to assume that the socio-economic situation of the border region contributed to the discrepancy between the two data sources.

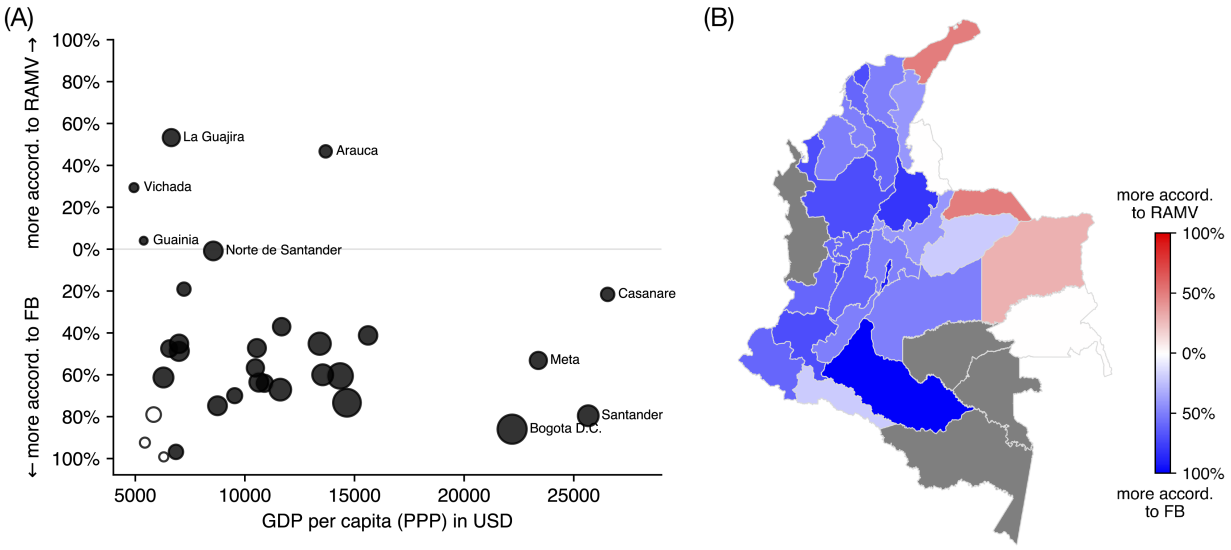


Figure 6: Bias in estimations of migrant and refugee numbers, comparing estimates for Colombia from Facebok and the RAMV survey. (A) Over- and underestimations of populations for individual departments in percentages. If Facebook estimates more refugees and migrants to be in one department we calculate the over-estimation compared to RAMV in percentage. If RAMV estimates more, we calculate the percentage compared to Facebook's estimate. The size of each point is scaled according to the population of the department. Departments with low numbers of migrants and refugees according to both Facebook and RAMV are colored white (1000 or below). (B) Departments adjacent to the Venezuelan border contain larger number of migrants and refugees than what is reported by Facebook. Departments for which both Facebook and RAMV have little data for are colored gray (1000 or below).

At a high level our work demonstrates the value that data held by private companies can have when used for public good [21], in particular in the domain of rapid disaster assessment [22]. We hope that our research helps to further the discussions on how to form Data Collaboratives [23] in support of humanitarian and development projects.

References

[1] Data available at r4v.info. Accessed on February 21, 2019.

[2] Emilio Zagheni and Ingmar Weber. You are where you e-mail: using e-mail data to estimate international migration rates. In *ACM Conference on Web Science*, pages 348–357, 2012.

[3] Bartosz Hawelka, Izabela Sitko, Euro Beinat, Stanislav Sobolevsky, Pavlos Kazakopoulos, and Carlo Ratti. Geo-located twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41(3):260–271, 2014.

[4] John A. Quinn, Marguerite M. Nyhan, Celia Navarro, Davide Coluccia, Lars Bromley, and Miguel Luengo-Oroz. Humanitarian applications of machine learning with remote-sensing data: review and case study in refugee settlement mapping. *Philosophical Transactions of the Royal Society A*, 376, 2018.

[5] Emilio Zagheni, Ingmar Weber, and Krishna Gummadi. Leveraging facebook's advertising platform to monitor stocks of migrants. *Population and Development Review*, 43(4):721–734, 2017.

[6] Report available at https://r4v.info/en/documents/details/68069. Accessed on February 24, 2019.

[7] Report available at https://www.iom.int/sites/default/files/press_release/file/rmrp_venezuela_2019_onlineversion_final.pdf. Accessed on February 24, 2019.

[8] Data available at https://data2.unhcr.org/en/documents/download/67311. Accessed on February 24, 2019.

[9] Nairar Galarraga Gortazar. Radiografía del gran éxodo venezolano. *El Pais*, 2018.

[10] Mónica Duarte. La nueva diáspora venezolana es cada vez más vulnerable y más masiva. *La Razon*, 2018.

[11] Juan Carlos Guataqui, Andres Garcia-Suaza, Cindy Vanessa Ospina-Cartagena, Diana Isabel Londono-Aguirre, Paul Rodriguez-Lesmes, and Juan Pablo Baquero. Informe 3: Caracteristicas de los migrantes de Venezuela a Colombia. *Universidad del Rosario*, 2017.

[12] Organizacion Internacional para las Migraciones. Monitoreo de flujo de poblacion venezolana: Chile. Technical report, International Organization for Migration, 2018.

[13] Pierre Deville, Catherine Linard, Samuel Martin, Marius Gilbert, Forrest R. Stevens, Andrea E. Gaughan, Vincent D. Blondel, and Andrew J. Tatem. Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences*, 111(45):15888–15893, 2014.

[14] David Pastor-Escuredo, Alfredo Morales-Guzmán, Yolanda Torres-Fernández, Jean-Martin Bauer, Amit Wadhwa, Carlos Castro-Correa, Liudmyla Romanoff, Jong Gun Lee, Alex Rutherford, Vanessa Frias-Martinez, et al. Flooding through the lens of mobile phone activity. In *IEEE Global Humanitarian Technology Conference (GHTC 2014)*, pages 279–286. IEEE, 2014.

[15] Alfredo J. Morales, Werner Creixell, Javier Borondo, Juan Carlos Losada, and Rosa Maria Benito. Characterizing ethnic interactions from human communication patterns in ivory coast. *Networks & Heterogeneous Media*, 10:87–99, 2015.

[16] Deepa K. Pindolia, Andres J. Garcia, Zhuojie Huang, Timothy Fik, David L. Smith, and Andrew J. Tatem. Quantifying cross-border movements and migrations for guiding the strategic planning of malaria control and elimination. *Malaria Journal*, 13(1):169, May 2014.

[17] David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. The parable of google flu: Traps in big data analysis. *Science*, 343(6176):1203–1205, 2014.

[18] Amaç Herdagdelen, Bogdan State, Lada A. Adamic, and Winter A. Mason. The social ties of immigrant communities in the united states. In *ACM Conference on Web Science*, pages 78–84, 2016.

[19] G. Venkatadri, A. Andreou, Y. Liu, A. Mislove, K. P. Gummadi, P. Loiseau, and O. Goga. Privacy risks with facebook's pii-based targeting: Auditing a data broker's advertising interface. In *IEEE Symposium on Security and Privacy (SP)*, pages 89–107, 2018.

[20] Amy Wesolowski, Nathan Eagle, Abdisalan M. Noor, Robert W. Snow, , and Caroline O. Buckee. The impact of biases in mobile phone ownership on estimates of human mobility. *Journal of The Royal Society Interface*, 10, 2013.

[21] Hetan Shah. Use our personal data for the common good. *Nature*, 556(7), 2018.

[22] Yury Kryvasheyeu, Haohui Chen, Nick Obradovich, Esteban Moro, Pascal Van Hentenryck, James Fowler, and Manuel Cebrian. Rapid assessment of disaster damage using social media activity. *Science Advances*, 2(3), 2016.

[23] Iryna Susha, Marijn Janssen, and Stefaan Verhulst. Data collaboratives as "bazaars"?: A review of coordination problems and mechanisms to match demand for data with supply. *Transforming Government: People, Process and Policy*, 11(1):157–172, 2017.

[24] Matheus Araujo, Yelena Mejova, Ingmar Weber, and Fabricio Benevenuto. Using facebook ads audiences for global lifestyle disease surveillance: Promises and limitations. In *Proceedings of the 2017 ACM on Web Science Conference*, pages 253–257. ACM, 2017.

[25] Data available at `https://esa.un.org/unpd/wpp/DVD/Files/1_Indicators\%20(Standard)/EXCEL_FILES/1_Population/WPP2017_POP_F01_1_TOTAL_POPULATION_BOTH_SEXES.xlsx`. Accessed on February 21, 2019.

Appendices

Materials and Methods

Estimating the Number of Refugees and Migrants from Venezuela

The dataset used in this paper was collected through the Facebook Marketing API (version 3.1). Fig. 7 shows a screenshot of the Web interface.[6] In this work, we automatically collected data using the Python library pySocialWatcher[7] [24]. This library provides a wrapper for the relevant calls in Facebook's Graph API[8].

Unless otherwise stated, the collection made in this work aim to capture Facebook monthly active users (MAUs) aged 13 or above which Facebook classified as belonging to the category *Expats - Venezuela*, which was renamed in late 2018 to *Lived in Venezuela (formerly Expats - Venezuela)*, as shown in Fig. 7. The definition of this category according to Facebook is *"People who used to live in Venezuela who now live abroad"*. Based on the validation results in the main article, we will also refer to these as estimates of Venezuelan refugees and migrants.

Correction Factor for Facebook Penetration in the Host Countries

When analyzing the temporal trends in Fig. 2, we included up-adjusted Facebook audience estimates that correct for the fact that not all refugees and migrants are on Facebook by taking into account the Facebook penetration in the host country. For that, we used the 2017 population estimations compiled by the United Nations [25]. An adjustment factor is calculated for each country as shown in Eq. 1. Estimates for refugees and migrants from Venezuela are then corrected as shown in Eq. 2.

$$Adj\_factor(country) = FB\_audience(country)/UN\_2017(country) \quad (1)$$

$$venezuelans\_adj(location) = venezuelans(country)/adj\_factor(country) \quad (2)$$

Table 1 shows both the raw Facebook estimates of refugees and migrants from Venezuela as well as adjusted values for different countries. The column *R4V* refers to the only official, but not necessarily accurate, data available from the Regional Inter-Agency Coordination Platform for Refugees and Migrants from Venezuela (R4V). To correct the estimates from June 2016, we used the 2016 population estimates from the United Nations instead of the 2017 one.

University Graduate Users in Facebook

The current Facebook Marketing API version has thirteen non-overlapping categories for education level.[9] In this work, the definition of "university graduate" combines the following five

[6] `https://business.facebook.com/adsmanager/creation/`

[7] `https://github.com/maraujo/pySocialWatcher`

[8] `https://developers.facebook.com/docs/marketing-apis`

[9] Details about the education levels supported by Facebook's Marketing API can be found at `https://developers.facebook.com/docs/marketing-api/targeting-specs/#education_and_workplace`, last accessed on February 28, 2019.

categories from Facebook: (1) *"At university (postgraduate)"*, (2) *"Doctorate degree"*, (3) *"Master's degree"*, (4) *"Some university (postgraduate)"*, (5) *"University graduate"*. Tab. 2 details information regarding university graduate Facebook users in different locations, both for the host population as well as for the refugees and migrants from Venezuela living in the same location. Part of the data shown in Tab. 2 can be found in Fig. 4A in the main article.

Linear Regression Model to Predict GDP

The prediction of the Gross Domestic Product (GDP) per capita at nominal values was conducted with an ordinary least squares linear regression model. We employed as ground-truth data to train the linear regression model the most recent GDP per capita data collection made in 2017 from the United Nation[10]. The linear regression model has only one independent variable X, representing the percentage of iOS-device users in the considered population. In detail, the targeting attribute used is called *"Facebook access (mobile): Apple (iOS) devices"* and is described as *"People who primarily access Facebook using an Apple (iOS) mobile device"*. The fitted model is $Y = 507.13 + X \cdot 104903.24$. The model, detailed in Table 3, reached a Mean Absolute Error of 3,782 and a Root Mean Squared Error of 4,537. Table 4 shows the raw data used in this experiment.

The model above, which predicts a country's GDP per capita at nominal values, is then applied to the sub-population of Facebook users who used to live in Venezuela and who now live in the different host countries and regions. The predictions for this sub-population can be found in Fig. 4B in the main article.

[10] Available Online at `https://unstats.un.org/unsd/snaama/Index`. Last accessed on February 11, 2019.

Table 1: Estimates of refugees and migrants from Venezuela in different countries used in Fig. 2.

Location	R4V		Facebook																			
	31 Oct 18	15 Jan 19	1 Jun 16	1 Jun 16 Adj	1 Jun 18	1 Jun 18 Adj	1 Jul 18	1 Jul 18 Adj	3 Aug 18	3 Aug 18 Adj	3 Sep 18	3 Sep 18 Adj	3 Oct 18	3 Oct 18 Adj	3 Nov 18	3 Nov 18 Adj	4 Dec 18	4 Dec 18 Adj	04 Jan 19	04 Jan 19 Adj	02 Feb 19	02 Feb 19 Adj
Argentina	130.0k	130.0k	24.0k	40.5k	100.0k	138.3k	110.0k	157.1k	110.0k	157.1k	120.0k	166.0k	120.0k	166.0k	130.0k	179.9k	130.0k	179.9k	140.0k	193.7k	140.0k	193.7k
Aruba	16.0k	16.0k	5.1k	7.6k	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	5.6k	11.6k	4.6k	10.1k	4.8k	6.2k	9.1k	10.8k
Bolivia	5.0k	5.0k	2.2k	6.7k	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	7.2k	12.1k	7.4k	12.0k	7.6k	12.2k	7.9k	12.5k
Brazil	75.0k	96.0k	12.0k	25.2k	53.0k	87.3k	56.0k	91.6k	52.0k	86.4k	60.0k	96.6k	67.0k	107.9k	73.0k	117.5k	75.0k	120.7k	79.0k	127.2k	83.0k	133.6k
Chile	108.0k	288.0k	28.0k	45.6k	220.0k	305.5k	220.0k	305.5k	220.0k	305.5k	240.0k	333.3k	250.0k	347.2k	260.0k	361.1k	280.0k	388.9k	290.0k	402.8k	310.0k	430.5k
Colombia	1.0M	1.1M	87.0k	192.4k	950.0k	1.5M	1.0M	1.6M	970.0k	1.5M	1.1M	1.7M	1.3M	2.0M	1.3M	2.0M	1.4M	2.1M	1.4M	2.1M	1.5M	2.2M
Costa Rica	25.0k	25.0k	5.9k	11.0k	15.0k	22.3k	15.0k	22.3k	15.0k	22.3k	15.0k	21.6k	16.0k	23.1k	16.0k	23.1k	16.0k	23.1k	17.0k	24.5k	17.0k	23.8k
Dominican Republic	28.5k	28.5k	12.0k	33.6k	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	52.0k	100.0k	52.0k	100.0k	52.0k	98.2k	53.0k	98.4k
Ecuador	221.0k	221.0k	20.0k	41.5k	150.0k	226.7k	160.0k	241.8k	160.0k	241.8k	180.0k	249.4k	200.0k	277.1k	220.0k	304.8k	230.0k	318.6k	240.0k	332.5k	250.0k	346.4k
Guyana	36.4k	36.4k	560.0	1.8k	3.9k	7.8k	4.1k	8.2k	3.6k	7.4k	4.1k	8.2k	4.7k	9.4k	5.0k	10.0k	5.4k	10.8k	5.5k	10.7k	5.8k	11.0k
Mexico	39.5k	39.5k	51.0k	110.2k	97.0k	150.9k	98.0k	152.5k	94.0k	146.3k	98.0k	150.7k	100.0k	152.0k	100.0k	155.6k	100.0k	152.0k	100.0k	150.2k	100.0k	148.5k
Panama	94.0k	94.0k	40.0k	107.6k	83.0k	179.0k	82.0k	176.9k	79.0k	179.9k	80.0k	172.6k	81.0k	174.7k	81.0k	174.7k	80.0k	172.6k	80.0k	172.6k	80.0k	172.6k
Peru	506.0k	506.0k	26.0k	55.1k	300.0k	419.5k	340.0k	475.5k	350.0k	511.7k	420.0k	587.4k	480.0k	671.3k	520.0k	727.2k	560.0k	783.2k	600.0k	804.1k	630.0k	844.3k
Paraguay	500.0	500.0	2.7k	7.6k	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	4.5k	8.8k	4.5k	8.5k	4.6k	8.7k	4.6k	8.5k
Trinidad and Tobago	40.0k	40.0k	2.4k	5.2k	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	11.0k	20.4k	12.0k	22.2k	12.0k	22.2k	12.0k	21.6k
Uruguay	8.5k	8.5k	2.5k	4.1k	8.3k	11.5k	8.6k	11.9k	8.5k	12.2k	9.1k	12.6k	9.5k	13.1k	9.9k	13.7k	10.0k	13.8k	11.0k	15.2k	11.0k	15.2k

Table 2: University graduate Facebook users. "University Graduate FB Users in host location" and "University Graduate FB Users from Venezuela" are the number of self-declared university graduate Facebook users from, respectively, the host population in a location and the migrants from Venezuela living in the same location. "% University Graduate Users in host location" and "% University Graduate Users from Venezuelans" are the percentage of self-declared university graduate Facebook users from, respectively, the host population in a location and the migrants from Venezuela living in the same location.

Location	University Graduate FB Users in host location	University Graduate FB Users from Venezuela	% University Graduate Users in host location	% University Graduate Users from Venezuelans
Argentina	9.1M	77.0k	30.7	54.6
Aruba	14.0k	3.3k	27.2	36.7
Bolivia	1.6M	3.5k	24.6	44.9
Brazil	35.0M	31.0k	27.6	37.8
Chile	3.5M	170.0k	29.9	54.1
Colombia	10.0M	530.0k	33.0	36.6
Costa Rica	850.0k	8.2k	27.7	49.4
Curacao	19.0k	3.4k	24.1	40.0
Dominican Republic	1.7M	25.0k	31.8	48.1
Ecuador	3.8M	110.0k	34.2	44.0
Spain	6.1M	130.0k	32.1	50.2
Guyana	91.0k	1.8k	24.3	31.6
Mexico	23.0M	47.0k	27.7	45.6
Panama	490.0k	38.0k	29.9	48.7
Peru	7.0M	290.0k	30.8	45.3
Trinidad and Tobago	250.0k	5.0k	37.9	40.7
United States	65.0M	210.0k	39.6	50.8
Uruguay	800.0k	6.0k	33.6	53.6
Venezuela	4.2M	-	35.9	-
Roraima, Brazil	82.0k	13.0k	28.3	34.9
São Paulo, Brazil	9.9M	4.0k	30.6	47.6
Miraflores, Peru	180.0k	9.6k	51.7	52.5

Table 3: Details of the linear model to estimate GDP.

Metric	**Value**
Model	OLS
Method	Least Squares
No. Observations	15
Df Residuals	13
Df Model	1
R-squared	0.883
Adj. R-squared	0.874
F-statistic	97.76
Prob (F-statistic)	2.05e-07
Log-Likelihood	-147.59
AIC	299.2
BIC	300.6
Omnibus	0.615
Prob(Omnibus)	0.735
Skew	0.013
Kurtosis:	2.034
Durbin-Watson	1.961
Jarque-Bera (JB):	0.583
Prob(JB)	0.747
Cond. No.	8.65

Table 4: Data used in the experiments to predict GDP per capita. "UN GDP'17" is the ground-truth estimations used by the linear model. "%iOS host population" and "%iOS Venezuelans" are the percentage of users that accessed Facebook with an iOS device from, respectively, the host population in a location and the refugees and migrants from Venezuela living in the same location. "GDP per capita predicted to host population" and "GDP per capita predicted to Venezelans" are the model GDP per capita predictions for both host population and Venezuelans in a location.

Location	UN GDP'17	%iOS host population	GDP per capita predicted to host population	%iOS Venezuelans	GDP per capita predicted to Venezelans
Argentina	14.4k	5.9	6.7k	15.3	16.6k
Aruba	25.7k	31.5	33.5k	18.8	20.2k
Brazil	9.8k	9.8	10.8k	4.6	5.4k
Chile	15.3k	14.6	15.9k	13.8	15.0k
Colombia	6.3k	6.8	7.6k	3.1	3.7k
Curacao	19.6k	22.6	24.2k	16.7	18.0k
Dominican Republic	7.1k	12.7	13.8k	17.5	18.9k
Ecuador	6.3k	7.3	8.2k	6.2	7.0k
Spain	28.4k	22.2	23.8k	28.5	30.4k
Mexico	9.0k	11.6	12.7k	23.2	24.9k
Panama	15.1k	8.3	9.2k	21.2	22.7k
Peru	6.6k	4.3	5.0k	4.2	5.0k
Trinidad and Tobago	16.1k	15.5	16.8k	10.8	11.8k
United States	60.1k	51.4	54.4k	54.3	57.5k
Uruguay	17.1k	13.0	14.2k	15.8	17.1k
Venezuela	-	4.4	5.1k	-	-
São Paulo, Brazil	-	12.6	13.7k	14.8	16.0k
Roraima, Brazil	-	7.5	8.4k	2.7	3.3k
Miraflores, Peru	-	16.7	18.0k	9.9	10.9k

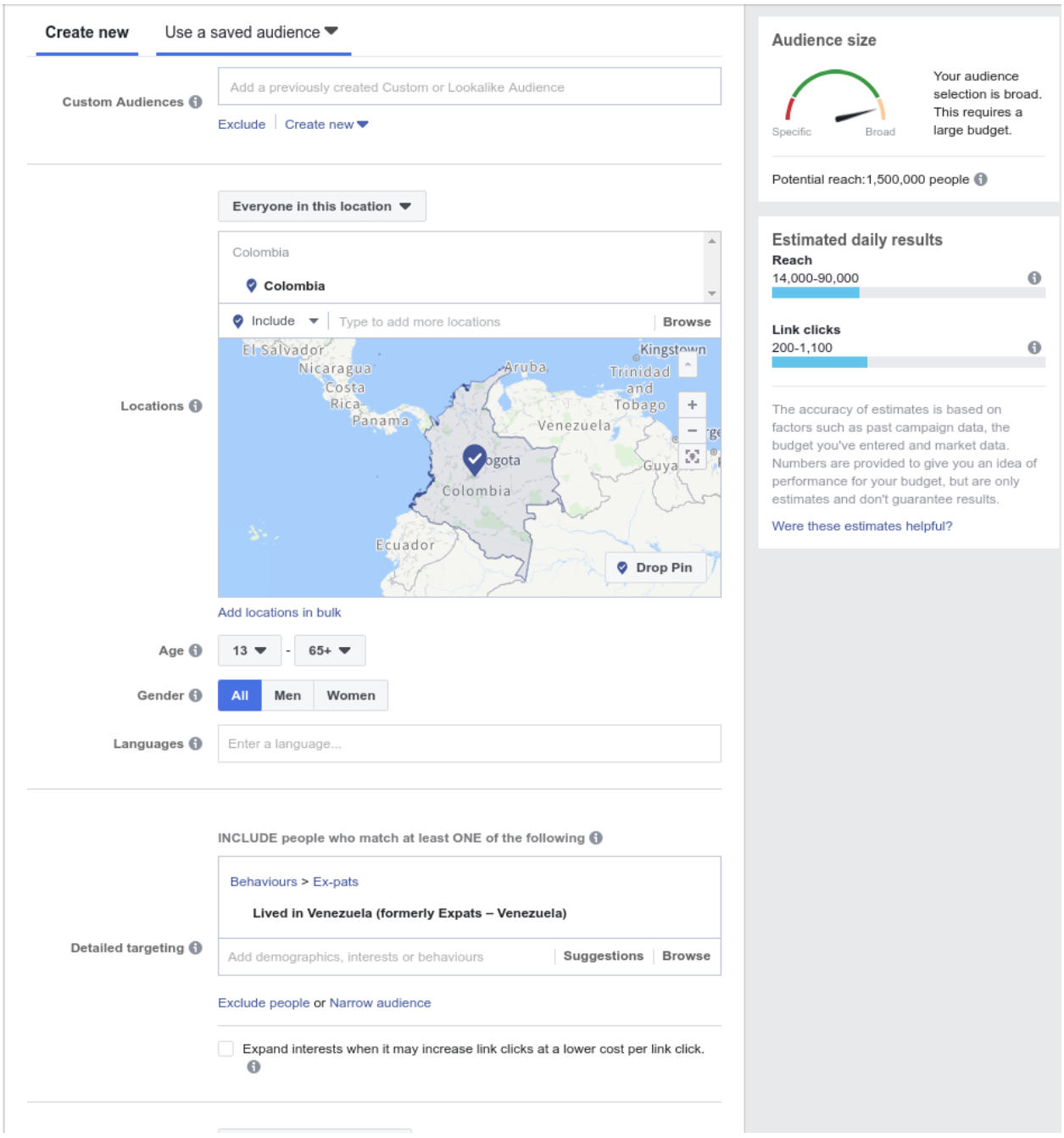


Figure 7: Screenshot of Facebook Marketing API's Web interface at `https://business.facebook.com/adsmanager/creation/`, taken on February 10, 2019. The Facebook Marketing API shows that an advertisement shown to Facebook users living in Colombia who previously lived in Venezuela can reach 1.5M monthly active users.