Explaining the travelling behaviour of migrants using Facebook audience estimates

Spyridon Spyratos*, Michele Vespe, Fabrizio Natale, Stefano Maria Iacus, Carlos Santamaria
European Commission, Joint Research Centre (JRC), Ispra, Italy

* Spyridon.SPYRATOS@ec.europa.eu

Abstract

The paper explores the travelling behaviour of migrant groups using Facebook audience estimates. Reduced geographical mobility is associated with increased risk of social exclusion and reduced socio-economic and psychological well-being. Facebook audience estimates are timely, openly available and cover most of the countries in the world. Facebook classifies its users based on multiple attributes such as the country of their previous residence, and whether they are frequent travellers. Using these data, we modelled the travelling behaviour of Facebook users grouped by countries of previous and current residence, gender and age. We found strong indications that the frequency of travelling is lower for Facebook users migrating from low-income countries and for women migrating from or living in countries with high gender inequality. Such mobility inequalities impede the smooth integration of migrants from low-income countries to new destinations and their well-being. Moreover, the reduced mobility of women who have lived or currently live in countries with conservative gender norms capture another aspect of the integration which is referring to socio-cultural norms and gender inequality. However, to provide more solid evidence on whether our findings are also valid for the general population, collaboration with Facebook is required to better understand how the data is being produced and pre-processed.

Introduction

This article aims to study and explain the travelling behaviour of migrant groups at a global level using Facebook audience estimates. The main idea behind studying the geographical mobility of migrants groups is that by understanding their travelling behavior it is possible to have an indirect measure of migrant’s integration and well-being. Several studies associate reduced geographical mobility with increased risk of social exclusion [1], reduced psychological well-being [2] and lower-income [3, 4]. De Vos et al. [5] suggests that travel behaviour affects well-being through experiences during a) destination-oriented travel; b) activity participation enabled by travel; c) activities during destination-oriented travel; d) trips where travel is the activity; and e) through potential travel which is defined by Kaufmann, Bergman, & Joye [6] as motility.
There is limited research focusing on the travelling behaviour of migrant communities. Shin [7] study demonstrated that the possibilities and potential for movement are often limited for women and minorities. Other studies [8, 9] demonstrated that the Russian speakers in Estonia are less spatially mobile within Estonia, but more mobile regarding travels outside Estonia compared to the Estonian speakers. According to Georggi and Pendyala [3] African Americans and Hispanics in the US have reduced long-distance travel behaviour compared to the white population.

Several studies relied on non-traditional data such as mobile position data, air traffic data, Twitter data and IP address data to study human mobility. Masso, Silm, & Ahas [9] studied domestic and international spatial mobility by age, gender and language (Estonian or Russian) using passive mobile positioning data from an Estonian mobile phone operator. Silm, Ahas, & Nuga [10] studied gender mobility difference using both mobile position data and questionnaire survey in Estonia. Hawelka et al. [11] used Twitter data to estimate the volume of international travellers by country of residence. Fiorio et al., [12] used Twitter data to study short-term mobility and long-term migration within the US. Gabrielli, Deutschmann, Natale, Recchi, & Vespe, [13] used monthly air passenger traffic to study types of mobility and mobility trends at a global level. Finally, State, Ingmar, & Zagheni, [14] study inter-national mobility using IP data recorded from the logins of an initial sample of over 100 million users of Yahoo Web service.

Most of these studies are either limited to specific groups of migrants, to specific countries, or they are capturing traffic flows between countries. Still, a systematic and comparative analysis of migrants’ mobility behaviour by country of origin and destination at a global level is missing. When it comes to statistical data sources, to the best of our knowledge, there are no available data sets at a global level about the domestic or international travelling behaviour which target specifically migrants. In the UK, the International Passenger Survey (IPS) [15] collects information about passengers departing from and arriving in the UK by nationality and residence among other attributes. At the European level, Eurostat [16] provides on annual basis statistics about domestic and outbound touristic trips of EU residents but also in this case the data are not designed to represent specifically migrants’ populations in each Member State.

Data from Facebook can help to address the absence of statistics about the mobility of migrant groups at a global level. Facebook, despite its limitations, offers unprecedented possibilities to capture new insights on the sociodemographic and behavioural characteristics of the migrants’ population broken down by country of residence, country of origin, age and gender. To the best of our knowledge, this is the first study to explore the travelling attribute of the Facebook Advertising platform [17]. By considering this additional attribute, we build on past research that makes use of data from Facebook to study the size of migrant stocks [18, 19], migrant assimilation [20] and gender inequalities [21]. The use of non-traditional data sources, like Facebook, can potentially complement traditional data as a source for statistics.

This article is organised as follows. The data section presents the Facebook audience estimates used in this study and includes some simple correlations to test the association between the travelling behaviour of Facebook users and travel data available from the UK International passenger survey. The methodology section describes six models used to explain how the travelling behaviour of Facebook users is affected by gender, age and countries of the previous residence of the migrants. In the results and discussion section, we present the outcome of the Facebook data analysis, and the conclusions are outlined in the final section.

Data

The Facebook Advertising platform, allows users to design targeted advertisements on the Facebook family of applications, by selecting the characteristics of the targeted audience.
These characteristics of the Facebook users include, for example, age, gender, location, country of previous residence and, particularly relevant for this study, whether they are “frequent travellers” or “frequent international travellers”. Once users have selected the characteristics of the Facebook population that they wish to target with the advertisement campaign, the advertising platform provides an estimate of the number of daily active users (DAU) and monthly active users (MAU) that fulfil these characteristics. We collected these estimates to generate aggregate estimates on the share of frequent travellers and frequent international travellers in the total population, by country of residence, age, and gender. We similarly collect the same estimates for the population that Facebook classifies as having lived abroad for all pairs of countries of previous and current destination as well as for the Facebook users who have not lived abroad.

Facebook classifies its users as “frequent international travellers” based on whether they have travelled abroad more than once in the past six months [17]. Since the data collection phase of this study took place from September 2019 to October 2019, it is expected that Facebook captured for the classification of the Facebook users as “frequent international travellers” international trips made up to six months before the date of data collection, meaning from March/April 2019 to September/October 2019. Facebook classifies its users as “frequent travellers” based on whether their activities on Facebook suggest that they are frequent travellers [17]. For the “frequent travellers” attribute Facebook does not provide any reference neither about the time period of the travel nor about the minimum distance of the travel required to classify its users as “frequent travellers”. The definition of frequent traveller is very generic since Facebook does not provide details on how it classifies a user as a “frequent traveller”. Thus we don’t know the destination of the travel, the purpose of the travel or whether the travels refer to short, middle or long-distance mobility. We decided to use the “frequent traveller” attribute in our analysis since we perform a comparative analysis of the same attribute between different “migrant” groups and the “non-migrant” population. The classification of Facebook users as “frequent international travellers” and “frequent travellers” is not likely to be self-reported. This is because, as of September 2019, the 51% of Facebook users who were mainly accessing Facebook through a mobile device were classified as frequent travellers, while only the 2.5% of those who were not primarily accessing Facebook using a mobile device were classified as frequent travellers. We can thus assume that Facebook is using the location of the mobile devices to classify users as frequent travellers or not.

To represent migrants, we rely on the classification of Facebook users as having “lived in country X”, which is based on whether they used to live in country X and they now live abroad. This classification is provided for the 89 countries of previous residence listed in S1 Table in Annex A. The key criteria that Facebook uses for identifying the previous residence of a user is the “hometown”, “current city”, and “other places lived”, as well the network structure of Facebook friendships [19]. In this study, we use the term Facebook “migrants” to describe Facebook users who have been classified as having lived in a country other than the country of their current residence and the term Facebook “non-migrants” to describe the users who have not lived in any other country than the country of their current residence.

To collect Facebook audience estimates, we have developed a python script, which was used to query the Facebook Marketing Application Programming Interface (API) [22] and store the data to a PostgresQL database. Using this python script, we collected for each age group $a \in [15–24, 25–34, 35–44, 45–54, 55–64, 15–64]$, gender $g \in \{Male, Female, Both\}$, country of current residence $c$, and country of previous residence $p \in \{countries \text{ in S1 Table}\}$, as well as, for non-migrants $n$ and total Facebook users $t$, the number of Facebook MAU $fb_{a,g,c,p/n/t}$; the number of Facebook MAU who are classified as “frequent international travellers” $fit_{a,g,c,p/n/t}$; and
the number of Facebook MAU who are classified as “frequent travellers” \( f_{\text{a, gccp}} \). We restricted the analysis only to Facebook users who primarily access Facebook using mobile devices since, as we explained earlier in this section, access from mobile devices represent a key feature for Facebook to classify the travelling behaviour of users. Due to the high number of variables collected and the API rate limits of approximately one API call every 10 seconds, the data collection period spanned from 4 September 2019 to 30 October 2019.

A first limitation of the collected Facebook audience estimates is that values are returned with a minimum threshold of 1000 “confidentiality threshold”. For example, if a selected group have 10 MAU, the Facebook estimate will be 1000 MAU. As a result, in this study, we are only able to use estimates about demographic groups with higher than 1000 MAU. A second limitation is that Facebook’s Marketing API only provides a rounded estimate of MAU. The applied rounding is proportional to the number of MAU, for example, for MAU values between 1000 and 10,000, the rounding precision is 100; for values between 10,000 and 100,000, the rounding precision is 1000; and so forth.

To assess the reliability of Facebook derived travelling estimates, we compared them with relevant statistics regarding international travels of UK residents. The International Passenger Survey (IPS) [15] collects information about passengers departing from and arriving in the UK by nationality and residence, among other attributes. Fig 1 shows the comparison between the log of the per capita number of international departures of UK residents by nationality during the time period March to August 2018 and the percentage of Facebook users who live in the UK and have at least made one international travel during the time period March to August 2019 by country of previous residence. To estimate the per capita number of international departure of UK residents by nationality, we divided the estimated number of international departures by the stock of UK migrants by citizenship available from Eurostat [22] for 2018. There is a good correlation between the two variables compared \( (R^2 = 0.6, p < 0.001) \), even though they differ in terms of reference time and definitions used to measure both

![Fig 1. International travelling behaviour in the UK by nationality/previous residence.](https://doi.org/10.1371/journal.pone.0238947.g001)
international travelling behaviour and country of birth or previous residence. The high $R^2$ is mainly due to countries with low values in both the x and y axes. Given the aim of this study, the below correlation for the case of UK shows that Facebook data can be used to identify, with a good degree of approximation, migrant groups with reduced international travelling behaviour such as Bangladeshi migrants.

A similar analysis has been carried out for the US. In this case, to the absence of national and international travelling statistics by country of origin, we decided to compare the frequent travelling Facebook attribute with income statistics. The income and the travelling frequency attribute do not measure the same phenomenon but the income explains part of the travelling behaviour. As Fig 2 shows, the estimated per capita annual income of individuals in US dollars for 2017 by country of birth in the US [23] is correlated ($R^2 = 0.46, p < 0.001$) with the percentage of frequent travellers in the US by country of previous residence.

**Methodology**

All data from Facebook’s Marketing API were provided to us in a fully anonymised, aggregated and rounded format with a confidentiality threshold of 1000 or more users. Thus this data can be considered to be ‘statistical data’ and not ‘personal data’. The mobility of Facebook users is expected to be affected by their demographic characteristics, such as age and gender and by the characteristics of the countries of their current residence and, in case these users are migrants, the characteristics of the country of the previous residence. We use six regression models to test the role of these variables in explaining the travelling behaviour of Facebook users considering both migrant and the non-migrant populations. The models $a$ and $b$ presented below explain the mobility of Facebook non-migrant users while the models $c, d, e,$ and $f$ refer to Facebook migrant users. The models $a, c,$ and $e$ explain the “frequent travellers” Facebook attribute while the models $b, d$ and $f$ explain the “frequent international travellers” attribute.
As the dependent variables in the above-presented models, we used the percentages of frequent international travellers fit_{ftp,c,p} and frequent travellers fit_{ft,c,p}. These percentages are estimated using Eq (1) and Eq (2) by dividing the number of Facebook MAU who are classified as “Frequent international travellers” fit_{a,g,c,p,n} or “Frequent travellers” fit_{a,g,c,p,n} of age a, gender g, country of residence c, and of country of previous residence p or n of non-migrants by the number of Facebook MAU fb_{a,g,c,p,n} of age a, gender g, country of residence c, and of country of previous residence p or n of non-migrants.

\[
\text{fit}_{\text{ftp},c,p} = \frac{\text{fit}_{a,g,c,p,n}}{\text{fb}_{a,g,c,p,n}} \quad (1)
\]

\[
\text{fit}_{\text{ft},c,p} = \frac{\text{fit}_{a,g,c,p,n}}{\text{fb}_{a,g,c,p,n}} \quad (2)
\]

In the models a, b, c and d we used the current-country specific fixed effects, which means that each country of residence has its coefficient, except the one country which is used as a reference. In the models e and f we used as independent variables the percentage of non-migrant Facebook users who are frequent travellers or frequent international travellers respectively by country of current residence, age and gender.

The per capita income in the country of the previous residence is expected to have a positive impact on the mobility of Facebook users. As income measure, we used the Gross Domestic Product (GDP) per capita GDPpc expressed in current United States (US) dollars available from the World Bank [24]. Gender inequality is another variable that could affect the travelling of female Facebook users. We used the Gender Development Index (GDI) available from the United Nations Development Programme [25]. The GDI reflects gender-based disparities in three dimensions, health, knowledge and living standards, and it is the ratio of female and male Human Development Index (HDI). GDI is equal to one when, women and men have the same HDI, above one when female fares better than male and below one in the opposite case. GDI is available for 164 countries. In our models for observations that describe male travelling behaviour, we fixed the gdi or the gdi values to 1.

The distance between the countries of current and previous residence of a Facebook user is expected to have an impact on the number of trips back to the country of previous residence. The trips of Facebook users to the country of previous residence accounts for a proportion of the total international travels since not all the international trips are towards the country of previous residence. In our models, we used the geodetic distances \text{dist}_{c,p} between countries of previous and current residence from the CEPII’s GeoDist dataset [26]. We selected to use the
“dist” variable of the CEPII’s GeoDist dataset which describes the geodesic distances between the most important cities/agglomerations of each country in terms of population.

Statistical analyses based on non-randomly selected samples of the population, such as the groups of the population who use Facebook through a mobile device, can lead to erroneous conclusions. A possible solution for correcting the selection bias of Facebook users would be the use of the Heckman correction [27]. However, since we rely on Facebook audience estimates on aggregated form and not on individual-level data, the use of Heckman correction is not feasible. To overcome this limitation, we assumed that the smaller the proportion of users who access Facebook mainly through a mobile device to the real population is, the higher is the probability that this sample will represent the most tech-savvy and wealthy part of the population which is more likely to be a frequent traveller and a frequent international traveller. The only exception to this hypothesis is the age group 15–25 where low Facebook use may be due to the use of alternative social media applications such as Instagram.

To estimate the selection bias due to the use of statistics that refer to a population who uses Facebook through a mobile device, we introduce as a variable to our models the penetration rate of the Facebook usage, \( \text{pen_rate}_{a,g,c,t} \). The penetration rate is estimated using Eq (3) by dividing the total number of Facebook users who access Facebook mainly through a mobile device which includes both Facebook migrants and non-migrant users \( \text{fb}_{a,g,c,t} \) of age, gender and country of residence by the population \( \text{UNDESA_pop}_{a,g,c} \) of age, gender and country of residence taken from the UNDESA statistics for the year 2019 (medium projection variant) [28]. We assume that cases of lower penetration rates for users of age 15–24 in respect of the age group 25–34, are determined by the use of other social media rather than by differences in technology adoption or wealth, in these cases we assigned the penetration rate of the age group 25–34 to the age group 15–24.

\[
\text{pen_rate}_{a,g,c,t} = \frac{\text{fb}_{a,g,c,t}}{\text{UNDESA_pop}_{a,g,c}}
\]

We fitted the above described six regression models using the Ordinary Least Squares (OLS) method, as well as using the Adaptive Elastic Net (AdaENet) method [29]. For the AdaENet, we used 80% of the observations as training data and the remaining 20% as testing data, and a 50-fold cross-validation for selecting the optimal lambda penalization parameter. We used a fixed alpha parameter equal to 0.5 to perform an equal combination of Ridge and Lasso regression. We implemented the AdaENet method using the "glmnet" package of the R software [30].

The decision of pairing OLS with AdaENet is because the second allows for contextual model selection (the Lasso contribution) and shrinkage (the Ridge contribution) estimation. Lasso tends to produce parsimonious models (by dropping some of the coefficients) which perform very well in predictions, while Ridge allows keeping in the model correlated coefficients, and this is very good for explaining the impact of group of variables on the outcome without necessarily dropping some of the coefficients. When OLS and AdaENet agree on the sign and amplitude of the coefficients, it is a good confirmation of the quality of the model in terms of descriptive and predictive power. AdaENet coefficients are usually smaller than the corresponding OLS ones but standard errors cannot be easily obtained. On the other hand, OLS makes it possible to evaluate the significance of those coefficients. For this reason, we present both evidence in Table 1 of the next section.

Before fitting our models, we filtered and cleaned the data that we used in the six proposed models. First, we took into consideration countries of residence with at least two available gender-age observations. Second, we did not take into consideration the age-gender-residence observations with Facebook low penetration rate \( \text{pen_rate}_{a,g,c,t} \) less than 10%. This threshold...
Explaining the travelling behaviour of migrants using Facebook audience estimates

Table 1. Regression models using the OLS method and the AdaENet method.

| Model description | Model a: Frequent travellers, Non-migrants | Model b: Frequent international travellers, Non-migrants | Model c: Frequent travellers, Migrants | Model d: Frequent international travellers, Migrants | Model e: Frequent travellers, Migrants, no country fixed effect | Model f: Frequent international travellers, Migrants, no country fixed effect |
|-------------------|-------------------------------------------|--------------------------------------------------------|---------------------------------------|--------------------------------------------------|-------------------------------------------------|-------------------------------------------------|
| dependent variables | $ft_{per^a\text{EN}}$ | $ft_{per^a\text{EN}}$ | $ft_{per^a\text{EN}}$ | $ft_{per^a\text{EN}}$ | $ft_{per^a\text{EN}}$ | $ft_{per^a\text{EN}}$ |
| Ordinary Least Squares | $B_{1s}$, Country of residence dummies | 127 out of 141 signifi. at $p < 0.1$ | 115 out of 141 signifi. at $p < 0.1$ | 92 out of 105 signifi. at $p < 0.1$ | 63 out of 99 signifi. at $p < 0.1$ |
| $B_{1s} ft_{per^a\text{EN}}$ | | | | | | 0.843*** |
| $B_{2s} ft_{per^a\text{EN}}$ | | | | | | 0.769*** |
| $B_{3s}$, age 25–34 (ref. 15–24) | -0.011*** | -0.015*** | -0.011*** | -0.011** | -0.005. | 0.007, $p = 0.13$ |
| $B_{4s}$, age 35–44 | -0.016*** | -0.022*** | -0.025*** | -0.022*** | -0.026*** | -0.016*** |
| $B_{5s}$, age 45–54 | -0.033*** | -0.037*** | -0.046*** | -0.052*** | -0.047*** | -0.052*** |
| $B_{6s}$, age 55–64 | -0.067*** | -0.064*** | -0.088*** | -0.083*** | -0.078*** | -0.088*** |
| $B_{7s}$, female (ref. male) | -0.018*** | -0.016*** | -0.011*** | -0.008. | 0.000, $p = 0.77$ | -0.002, $p = 0.58$ |
| $B_{8s}$, (1—gd1i) | -0.258*** | -0.224*** | -0.281*** | -0.525*** | | |
| $B_{9s}$, (1—gd2i) | -0.208*** | -0.254*** | -0.235*** | -0.302*** | | |
| $B_{10s}$, log (dist cj) | -0.017 *** | -0.030 *** | -0.026 *** | -0.040 *** | | |
| $B_{11s}$, log (GDPcr) | 0.012*** | 0.016*** | 0.016*** | 0.020*** | | |
| $B_{12s}$, pen_rate a,g,pt | -0.021 | -0.030** | -0.039*** | -0.056** | -0.075*** | -0.170*** |
| Intercept | 0.608*** | 0.235*** | 0.690*** | 0.551*** | 0.269*** | 0.412*** |
| Multiple $R^2$ | 0.987 | 0.952 | 0.873 | 0.752 | 0.804 | 0.643 |
| degrees of freed. | 1023 | 1014 | 5649 | 2937 | 5848 | 3100 |
| Adaptive Elastic Net | | | | | | |
| $B_{1s}$, Country of residence dummies | 141 included out of 141 | 140 included out of 141 | 105 included out of 105 | 92 included out of 93 | | |
| $B_{1s} ft_{per^a\text{EN}}$ | | | | | | 0.839 |
| $B_{2s} ft_{per^a\text{EN}}$ | | | | | | 0.767 |
| $B_{3s}$, age 25–34 (ref. 15–24) | -0.007 | -0.013 | -0.010 | -0.008 | -0.0034 | 0.006 |
| $B_{4s}$, age 35–44 | -0.012 | -0.022 | -0.023 | -0.015 | -0.023 | -0.011 |
| $B_{5s}$, age 45–54 | -0.027 | -0.035 | -0.044 | -0.050 | -0.045 | -0.048 |
| $B_{6s}$, age 55–64 | -0.060 | -0.063 | -0.084 | -0.079 | -0.071 | -0.082 |
| $B_{7s}$, female (ref. male) | -0.018 | -0.017 | -0.011 | -0.008 | 0.0003 | -0.002 |
| $B_{8s}$, (1—gd1i) | -0.258 | -0.230 | -0.278 | -0.533 | | |
| $B_{9s}$, (1—gd2i) | -0.211 | -0.270 | -0.226 | -0.334 | | |
| $B_{10s}$, log (dist cj) | -0.017 | -0.030 | -0.026 | -0.040 | | |
| $B_{11s}$, log (GDPcr) | 0.012 | 0.018 | 0.016 | 0.019 | | |
| $B_{12s}$, pen_rate a,g,pt | -0.016 | -0.027 | -0.029 | -0.046 | -0.0867 | -0.160 |
| Intercept | 0.471 | 0.211 | 0.738 | 0.48 | 0.2533 | 0.423 |
| Multiple $R^2$ | 0.979 | 0.929 | 0.865 | 0.772 | 0.805 | 0.645 |
| No. of observation | 1172 | 1163 | 5765 | 3047 | 5859 | 3111 |

$p < 0.1$
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001.$

https://doi.org/10.1371/journal.pone.0238947.t002
is indent to exclude countries where Facebook is not the popular social media application like Russia and Uzbekistan as well as age-gender-residence observations with a low proportion of users who access Facebook mainly through a mobile device to the real population, to avoid bias linked to a very poor representation of Facebook in the overall population of the country. Finally, we excluded two countries out of 89 countries of previous residence listed in the S1 Table. China was excluded from the analysis since Facebook use is restricted in that country. Greece was also excluded as a country of previous residence since Facebook audience estimates for users who have lived in Greece are strongly underestimated most likely due to a Facebook classification error. As of December 2019, Facebook was reporting only 3,800 users who have lived in Greece and now live abroad while according to UNDESA [31] in 2017 there were 993,000 Greek-born citizens that live abroad.

Results

The main contribution of this research is that we found strong indications that the frequency of travelling is lower for Facebook users migrating from low-income countries and for women migrating from countries with high gender inequality. In Table 1, we present the unstandardized coefficients as well as the accuracies of the proposed six models using the OLS method and the AdaENet method. In Fig 3 we present the importance of the variables which were used in each of the six models. The variable importance was estimated using the impurity (Gini) importance of the Random Forest classifier available in package “ranger” of the R software [32]. This importance measure summarizes how frequently a variable is determinant in predicting the outcome variable in a Random Forest.

In line with the literature [9, 10, 33, 34], the models’ unstandardized coefficients reported in Table 1 show that women travel slightly less than men and the elderly travel less than the young. As shown in Fig 3 the GDI plays a much more important role in explaining gender inequalities in travelling behaviour that the gender categorical variable. The GDI of both the country of previous and current residence $gdi_p$ and $gdi_c$ is correlated with the general as well as international travelling mobility of Facebook female users. This means that female Facebook users who have lived or currently live in countries with conservative gender norms are travelling less compared to female Facebook users who live or have lived in countries where both genders fare equally. This relation is pointing to the possibility of capturing through the analysis of mobility patterns another aspect of the integration of migrants which is pertaining to socio-cultural norms and gender inequality.

The per capita GDP of the country of previous residence of a Facebook migrant user $GDP_{pc_p}$ is positively correlated with general as well as international travelling mobility. This positive effect and the importance of this variable corroborates the main idea that mobility patterns may offer an indication of the wealth of migrants.

As shown in Fig 3, the distance $dist_{c,p}$ between the country of current and previous residence of a Facebook user, plays an important role in their travelling behaviour. The $B_d$ unstandardized coefficient in all the four models, is negative, and it is higher and more important in the models $d$ and $f$ that explicitly describe the international travelling behaviour of Facebook migrant users. This is because part of the international trips is expected to have as destination the country of the previous residence. The inclusion of this variable in the models is important to neutralise the impact of the cost of reaching the home-country on migrants mobility.

The Facebook penetration rate $pen_{rate_{a,g,c}}$ is negatively correlated with the percentage of frequent or frequent international travellers in all the models. This variable is an important element in our model to compensate for the bias introduced by the over-representation in
Facebook of the most tech-savvy and wealthy part of the population, which is also more likely to include frequent travellers and frequent international travellers.

Finally, we provide a descriptive representation of the two Facebook attributes, namely the ‘frequent traveller’ in Fig 4 and of the ‘frequent international traveller’ in Fig 5. Apart from explaining the travelling behaviour of Facebook migrants based on the income and gender inequalities, we also identified migrant groups which according to Facebook audience estimates have very limited mobility, such as the Ethiopians and the Bangladeshis in Bahrain and Kuwait. This low mobility might represent domestic workers who are recruited through the Kafala system. Kafala system is a government policy used to organise and control the migrant workers in the Gulf Cooperation Council countries [35]. Finally, we also find that Facebook users who have lived in east European countries and now live in west European countries are
Fig 4. The heatmap of frequent travellers. The difference in the percentage of frequent travellers between Facebook migrant users and Facebook non-migrant users of age 15 to 64 for each country of residence. Empty cells mean no available data. The heatmap includes the 60 countries of residence with the highest number of migrant groups and the 50 countries of previous residence with the highest presence in the selected 60 countries of residence.

https://doi.org/10.1371/journal.pone.0238947.g004
Fig 5. The heatmap of frequent international travellers. The difference in the percentages of frequent international travellers between Facebook migrant users and Facebook non-migrant users of age 15 to 64 for each country of residence. Empty cells mean no available data. The heatmap includes the 60 countries of residence with the highest number of migrant groups and the 50 countries of previous residence with the highest presence in the selected 60 countries of residence.

https://doi.org/10.1371/journal.pone.0238947.g005
very mobile. However, we do not know to what percentage these Facebook users represent cross-border seasonal workers or permanent migrants.

Discussion and conclusions
The main objective of this paper has been to examine the travelling behaviour of different migrant groups in multiple countries using Facebook audience estimates. Based on Facebook audience estimates we found strong indications that Facebook migrant users who have lived in low-income countries are less mobile than Facebook migrants from rich countries. We also found that female Facebook users who have lived or currently live in countries where gender inequality is high are less mobile than female Facebook users who have lived or live in more gender-equal countries. We were able to identify Facebook migrant users with reduced travelling behaviour, such as Facebook users who have lived in Ethiopia, and now live in the Gulf countries.

There are various limitations related to the use of Facebook audience estimates. A first limitation is that Facebook penetration rates vary based on the users’ age, gender, origin, income, educational attainment, and on whether they live in urban or rural areas [19, 36, 37]. Clearly, Facebook users do not represent the real population, and thus to reduce the impact of the Facebook usage selection bias we introduced the penetration rate as a variable in our models. A second limitation is that the Facebook estimates are accessed in an opportunistic manner. Facebook may change at any time the conditions for accessing the data, it does not disclose the detailed criteria for classifying its users e.g. as a “frequent travellers”, and the classification criteria may change at any time without prior notice [19, 37].

A third limitation, but also a data protection safeguard, is that we have access to anonymized, aggregated and rounded data with a confidentiality threshold of 1000 users. On the one hand, due to the aggregated form of the data, we were not able to apply a more robust bias correction methodology, and due to the 1000 users confidentiality threshold, we did not obtain estimates about age-gender-residence-previous residence groups with less than 1000 users. On the other hand, the aggregated form of the data and the 1000 users confidentiality threshold guarantees that the re-identification of individuals is highly unlikely. Still, the high confidentiality threshold cannot eliminate the risk of exposing data about the location and the behaviour of large vulnerable migrant groups (e.g. displaced populations) when data are collected at a very detailed spatial-temporal resolution. Thus, as also concluded by Rama et al. [37], the use of Facebook audience estimates should be done with caution, especially in high-risk settings, for example in or near conflict zones.

Based on the literature reduced geographical mobility is associated with increased risk of social exclusion [1], reduced psychological well-being [2] and lower-income [3, 4]. When travelling is limited and devoted mostly to compulsory places, the whole experience of space becomes ruled by the sign of necessity, a space of survival rather than a space of belonging [38]. The importance of this study is to offer a novel possibility to build indicators of migrants’ well-being by measuring their geographical mobility. While such indicators can be constructed for specific groups of migrants, it becomes extremely challenging to have a more comprehensive and systematic overview at a global level. Facebook, despite its limitations, offers unprecedented possibilities to generate new statistics on the sociodemographic and behavioural characteristics of the migrants’ population broken down by country of residence, country of origin age and gender. In our study, by analyzing the different mobility patterns of migrants groups, we were able to show how mobility inequalities in the countries of previous residence of Facebook users are being perpetuated in the new countries of residence, a fact that can introduce structural barriers in the smooth integration of migrants, for example of women
from countries with conservative gender norms in western societies. However, in order to provide more solid evidence on whether our findings are also valid for the general population, collaboration with Facebook is required to better understand how the data is being produced and pre-processed.

Supporting information
S1 Table. List of countries of previous residence for which Facebook provides audience estimates.

(DOCX)

Acknowledgments
We would like to thank our colleagues at the European Commission’s Knowledge Centre on Migration and Demography for their valuable comments and suggestions.

Author Contributions
Conceptualization: Spyridon Spyratos, Michele Vespe, Fabrizio Natale.
Data curation: Spyridon Spyratos.
Formal analysis: Spyridon Spyratos, Stefano Maria Iacus.
Investigation: Spyridon Spyratos, Michele Vespe, Fabrizio Natale.
Methodology: Spyridon Spyratos, Michele Vespe, Fabrizio Natale, Stefano Maria Iacus, Carlos Santamaria.
Project administration: Spyridon Spyratos, Michele Vespe.
Software: Spyridon Spyratos.
Supervision: Michele Vespe.
Validation: Spyridon Spyratos.
Visualization: Spyridon Spyratos.
Writing – original draft: Spyridon Spyratos.
Writing – review & editing: Spyridon Spyratos, Michele Vespe, Fabrizio Natale, Stefano Maria Iacus, Carlos Santamaria.

References
1. Stanley JK, Hensher DA, Stanley JR, Vella-brodrick D. Mobility, social exclusion and well-being: Exploring the links. Transp Res Part A. 2011; 45: 789–801. https://doi.org/10.1016/j.tra.2011.06.007
2. Vella-brodrick DA, Stanley J. The significance of transport mobility in predicting well-being. Transp Policy. 2013; 29: 236–242. https://doi.org/10.1016/j.tranpol.2013.06.005
3. Georggi NL, Pendyala RM. Analysis of Long-Distance Travel Behavior of the Elderly and Low Income. Transportation Research—Personal Travel The Long and Short of It. Washington DC; 2001. ISSN 0097-8515
4. Gauvin L, Tizzoni M, Piaggese S, Young A, Adler N, Verhulst S, et al. Gender gaps in urban mobility. Humanit Soc Sci Commun. 2020; 7. https://doi.org/10.1057/s41599-020-0500-x
5. De Vos J, Schwanen T, van Acker V, Witlox F. Travel and Subjective Well-Being: A Focus on Findings, Methods and Future Research Needs. Transp Rev. 2013; 33: 421–442. https://doi.org/10.1080/01441647.2013.815665
6. Kaufmann V, Bergman MM, Joyce D. Motility: Mobility as Capital. Int J Urban Reg Res. 2004; 28: 745–756.
7. Shin H. Spatial Capability for Understanding Gendered Mobility for Korean Christian Immigrant Women in Los Angeles. Urban. 2011; 48: 2355–2373. https://doi.org/10.1177/0042098010388955

8. Järv O, Mürisepp K, Ahas R, Derudder B, Willox F. Ethnic differences in activity spaces as a characteristic of segregation: A study based on mobile phone usage in Tallinn, Estonia. Urban Stud. 2015; 52: 2680–2698. https://doi.org/10.1177/0042098015545049

9. Masso A, Silm S, Ahas R. Generational differences in spatial mobility: A study with mobile phone data. Popul Space Place. 2019; 25: 1–15. https://doi.org/10.1002/psp.2210

10. Silm S, Ahas R, Nuga M. Gender differences in space-time mobility patterns in a postcommunist city: A case study based on mobile positioning in the suburbs of Tallinn. Environ Plan B Plan Des. 2013; 40: 814–828. https://doi.org/10.1068/b38068

11. Hawelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopoulos P, Ratti C. Geo-located Twitter as proxy for global mobility patterns. Cartogr Geogr Inf Sci. 2014; 41: 260–271. https://doi.org/10.1080/15230406.2014.890072

12. Fiorio L, Abel G, Cai J, Zaghini E, Weber I, Vinué G. Using Twitter Data to Estimate the Relationships between Short-term Mobility and Long-term Migration. Proceedings of the 2017 ACM on Web Science Conference. ACM; 2017. pp. 103–110.

13. Gabrielli L, Deutschmann E, Natale F, Recchi E, Vespe M. Dissecting global air traffic data to discern different types and trends of transnational human mobility. EPJ Data Sci. 2019; 8. https://doi.org/10.1140/epjds/s13688-019-0204-x

14. State B, Ingmar W, Zaghini E. Studying Inter-National Mobility through IP Geolocation. WSDM’13. Rome, Italy: ACM; 2013. pp. 265–274.

15. Office for National Statistics. International Passenger Survey. In: UK Data Service [Internet]. 2019 [cited 4 Sep 2019]. Available: http://doi.org/10.5255/UKDA-5936-5

16. Eurostat. Annual data on trips of EU residents. 2019 [cited 27 Sep 2019]. Available: https://ec.europa.eu/eurostat/cache/metadata/en/tour_dem_esms.htm

17. Facebook. Facebook Ads Manager. 2020 [cited 15 Mar 2020]. Available: https://www.facebook.com/"admanager"/creation

18. Zaghini E, Weber I, Gummiadi K. Leveraging Facebook’s Advertising Platform to Monitor Stocks of Migrants. Popul Dev Rev. 2017; 43: 721–734.

19. Spyratos S, Vespe M, Natale F, Weber I, Zaghini E, Rango M. Quantifying international human mobility patterns using Facebook Network data. PLoS One. 2019; 14. https://doi.org/10.1371/journal.pone.0224134

20. Dubois A, Zaghini E, Garimella K, Weber I. Studying Migrant Assimilation Through Facebook Interests. In: Staab S, Kolitsa O, Ignatov D, editors. Social Informatics SocInfo 2018 Lecture Notes in Computer Science. Cham: Springer; 2018. pp. 51–60. Available: http://arxiv.org/abs/1801.09430

21. Fatěhkia M, Kashyap R, Weber I. Using Facebook ad data to track the global digital gender gap. World Dev. 2018, 107: 189–209. https://doi.org/10.1016/j.worlddev.2018.03.007

22. Eurostat. Population on 1 January by age group, sex and citizenship. 2018. Available: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=migr_pop&lang=en

23. American Community Survey. S0201—Selected population profile in the United States. 2017 [cited 14 Oct 2019]. Available: https://www.census.gov/

24. World Bank. GDP per capita (current US$). 2019 [cited 1 Oct 2019]. Available: https://data.worldbank.org/indicator/NY.GDP.PCAP.PC.CD

25. UNDP. Gender Development Index (GDI). 2017 [cited 30 Sep 2019]. Available: https://hdr.undp.org/en/content/gender-development-index-gdi

26. Mayer T, Zignago S. Notes on CEPII’s distances measures: The GeoDist database. 2011. Available: http://www.cepii.fr/anglaisgraph/bdd/distances.htm

27. Puhani PA. The Heckman Correction for Sample Selection and Its Critique. J Econ Surv. 2000; 14: 53–68.

28. UNDESA. Population by 5-year age groups, annually from 1950 to 2100: medium projection variant. 2017 [cited 20 Mar 2018]. Available: https://esa.un.org/unpd/wpp/Download/Standard/CSV/

29. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol. 2005; 67: 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

30. Friedman J, Hastie T, Tibshirani R, Simon N, Narasimhan B, Qian J. Package ‘glmnet.’ 2019. Available: https://cran.r-project.org/web/packages/glmnet/glmnet.pdf

31. UNDESA. Trends in International Migrant Stock: The 2017 Revision (United Nations database, POP/DB/MIG/Stock/Rev.2017). United Nations; 2017 p. 16. Available: http://www.un.org/en/development/desa/population/migration/data/estimates2/estimates17.shtml
32. Wright MN, Ziegler A. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. J Stat Softw. 2017; 77: 1–17. https://doi.org/10.18637/jss.v077.i01

33. Crane R. Revolution in Women’s Gender Gap in Commuting. J Am Plan Assoc. 2014; 73: 298–316. https://doi.org/10.1080/01944360708977979

34. Dargay JM, Clark S. The determinants of long distance travel in Great Britain. Transp Res Part A. 2012; 46: 576–587. https://doi.org/10.1016/j.tra.2011.11.016

35. Malit FT, Naufal G. Asymmetric Information under the Kafala Sponsorship System: Impacts on Foreign Domestic Workers’ Income and Employment Status in the GCC Countries. Int Migr. 2016; 54: 76–90. https://doi.org/10.1111/imig.12269

36. Smith A, Anderson M. Social Media Use in 2018. Pew Research Center; 2018. Available: http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/

37. Rama D, Mejova Y, Tizzoni M, Kalimeri K, Weber I. Facebook Ads as a Demographic Tool to Measure the Urban-Rural Divide. Web Conf 2020—Proc World Wide Web Conf WWW 2020. 2020; 327–338. https://doi.org/10.1145/3366423.3380118

38. Ureta S. To Move or Not to Move? Social Exclusion, Accessibility and Daily Mobility among the Low-income Population in Santiago, Chile. Mobilities. 2008; 3: 269–289. https://doi.org/10.1080/1745010802095338