Examining Global Mobile Diffusion and Mobile Gender Gaps through Facebook's Advertising Data

Nazanin Sabri
University of Tehran
Tehran, Iran
nazanin.sabri@ut.ac.ir

Ridhi Kashyap
University of Oxford
Oxford, United Kingdom
ridhi.kashyap@nuffield.ox.ac.uk

Ingmar Weber
Qatar Computing Research Institute
Doha, Qatar
iweber@hbku.edu.qa

ABSTRACT

Social media advertising data, particularly data from Facebook's advertising platform, have been successfully used for monitoring population and development indicators, with an emphasis on monitoring digital gender inequality. This paper contributes to this literature by assessing the feasibility of the attribute of user behavior of "using a mobile device for X months" available from Facebook's advertising platform to understand short-term global mobile diffusion dynamics and mobile phone gender gaps. We compare this attribute with other features of the platform to form a better understanding of the data and the digital behaviours they capture, and show how this Facebook attribute relates to mobile phone penetration rates and gender gaps in mobile access. We find that this "Uses a mobile device (X months)" advertising targeting attribute can be used as a proxy for changes in mobile phone penetration rates, especially among younger users, and that it captures cross-national variation in mobile gender gaps. We further find that countries with larger gender gaps disfavoring women are comparatively more gender-equal among the most recently joined cohort.

CCS CONCEPTS

• **Human-centered computing** → *Collaborative and social computing*; • **Applied computing** → Sociology; • **Social and professional topics** → **User characteristics**; • **Information systems** → **Online advertising**.

KEYWORDS

Gender Gaps, Access to mobile phones, Facebook's Advertisement Platform, Digital Demography, Online Advertising

ACM Reference Format:
Nazanin Sabri, Ridhi Kashyap, and Ingmar Weber. 2021. Examining Global Mobile Diffusion and Mobile Gender Gaps through Facebook's Advertising Data. In *Proceedings of the 32nd ACM Conference on Hypertext and Social Media (HT '21), August 30-September 2, 2021, Virtual Event, Ireland.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3465336.3475120

HT '21, August 30-September 2, 2021, Virtual Event, Ireland

ACM ISBN 978-1-4503-8551-0/21/08…$15.00
https://doi.org/10.1145/3465336.3475120

1 INTRODUCTION AND BACKGROUND

Social media advertising data have been increasingly used as a novel source of data within the broader genre of work exploring the potential for "big data" sources to model sustainable development indicators [18]. These data provide aggregate estimates of the size of the potential audiences or users of a given online platform who match specified demographic or user characteristics. Facebook's advertising platform has been the most widely used among different platforms due to the large size of the user population as the most widely-used social media site, counting 2.7 billions users[1] as of January 2021 [17], global coverage across different countries, and the ready and no-cost availability of ad audience estimates through the marketing API[2]. Due to these features, data from Facebook's marketing API have been used for modelling demographic differentials [9, 14], and development indicators linked to global internet and mobile phone access gender gaps [6, 11], migration [1, 19], fertility [13], poverty [5, 7], and population health [2].

A growing body of work has also used Facebook advertising data to study global digital gender gaps. Digital gender inequalities are an important dimension of population inequality, and have been recognised as such in the context of the global sustainable development goals (SDGs) [3]. In this context, [6] and [11] have highlighted the value of gender gaps on the platform for predicting country-level internet and mobile access gender gaps across the world, and expanding geographical coverage of digital gender gap indicators to countries beyond those for which traditional survey data are available. In particular, these studies show that where ground truth indicators from survey data sources exist, Facebook gender gaps show strong correlation with internet and mobile access gender gaps [6], as well as low-level digital skills gender gaps [11].

Although work using Facebook advertising data for modelling social phenomena has grown, most studies using these data draw on data collected at one time point, which provides a snapshot of the current user population, to model a given social indicator. A key shortcoming of Facebook's advertising data, but also more generally of all advertising platforms, is their inability to provide historic estimates of past user counts. This makes it difficult to track changes over time unless the data are routinely collected for repeated snapshots. Furthermore, potential changes over time in the underlying algorithm providing audience counts can often make the interpretation of repeated snapshots of data collection challenging. Nevertheless, understanding changes over time in this social media population has significant value for expanding the

[1] There are different methodologies for defining and counting "users". Facebook's own advertising platform provides a lower estimate of 2.2 billion monthly active users (MAUs) as of April 2021.
[2] https://developers.facebook.com/docs/marketing-apis/
[3] https://sdg.tracking-progress.org/indicator/5-b-1-mobile-telephone-ownership/

potential applications of these data, for example by modelling temporal changes in social indicators of interest.

In this study, we explore the value of specific attributes within Facebook's advertising platform on users' duration of access to mobile phones to understand dynamics over time of this online population. As a number of studies have specifically examined digital gender gaps using Facebook data [6, 8, 11, 12], to examine if this attribute enables us to capture historic snapshots of Facebook's user population, we also focus on the potential offered through the lens of gender gap indicators. We further explore whether this attribute can provide us with insight into other digital behaviours, particularly those where data gaps in traditional data sources such as surveys exist. For example, while these data directly indicate changes linked to the Facebook user population, they could also offer insights into changes in the global digitally connected population, such as, changes in the demographic characteristics of mobile users (e.g. gender gaps of mobile users) or provide indirect indicators of multiple device use or connectivity behaviours, both of which are digital behaviours not routinely captured in traditional data sources or global telecommunication statistics. We assess these different perspectives using a cross-country analysis of the user duration attribute, along with a validation check with external data.

2 METHODS AND MATERIAL

For our study, we have used Facebook's Advertising platform [4]. Facebook offers a wide range of advertising targeting features, including whether someone "uses a mobile device (T months)". Here T can have any of the following values: less than 1 month, 1-3 months, 4-6 months, 7-9 months, 10-12 months, 13-18 months, 19-24 months, 25 months+. Facebook describes these features as "people who are likely to have used a mobile device for T months". This targeting attribute is the primary feature that we analyze in this study.

Initially we collected Facebook data for 247 countries, territories, or areas of geographical interest that have assigned official codes in ISO 3166-1 and are supported by Facebook. However, to avoid sparsity and to be consistent with prior work [6, 11] we limit our country-level analysis to the 168 countries with > 200k monthly active Facebook users. We further removed China from our calculations as access to Facebook is restricted in China and thus the audience estimates collected from the platform are likely for users with VPN access, who might not reflect the mobile phone usage in the country more generally.

3 RESULTS

The description of the targeting attribute "People who are likely to have used a mobile device for X months" is somewhat ambiguous, and Facebook and other advertising platforms do not provide public methodological details or detailed specifications. To better understand this attribute, we first perform a validation exercise to understand how to best interpret these audience counts. Next we examine gender gaps as derived from the "known duration of mobile device usage" feature.

3.1 Understanding and Ground-Truthing the Targeting Attribute

As of April 2021, Facebook's advertising platform reports an estimated 2.2 billion monthly active users worldwide who access Facebook through a mobile device. Of these, 1.5 billion (68%) have a known duration of using a mobile phone. Based on available statistics, we know that 98.3% of Facebook users access Facebook through a mobile phone, though a number of these users additionally use a desktop or tablet device to access Facebook [16].

Our first hypothesis for the device usage duration information is that it is tied to users who *primarily use a single device* to access Facebook. Support for this is provided by the fact that other behavior based targeting attributes, such as "Facebook access (mobile): Android devices" or "Facebook access (mobile): iOS devices" all refer to "primarily" in their description. Additionally, the sum of users who primarily use either Android or iOS, as estimated using the behavior feature, adds up to only 1.7B whereas the operating system targeting adds up to 2.2B. This potentially indicates that at least 500M Facebook users might occasionally use a mobile device that's not their primary device. Further, Figure 1a shows that the proportion of users with known duration of mobile phone usage among Facebook users is higher in low-income countries (e.g. in regions of Sub-Saharan Africa and South Asia), and lower in high-income countries (e.g. North America, Western Europe) where individuals are more likely to own multiple devices.

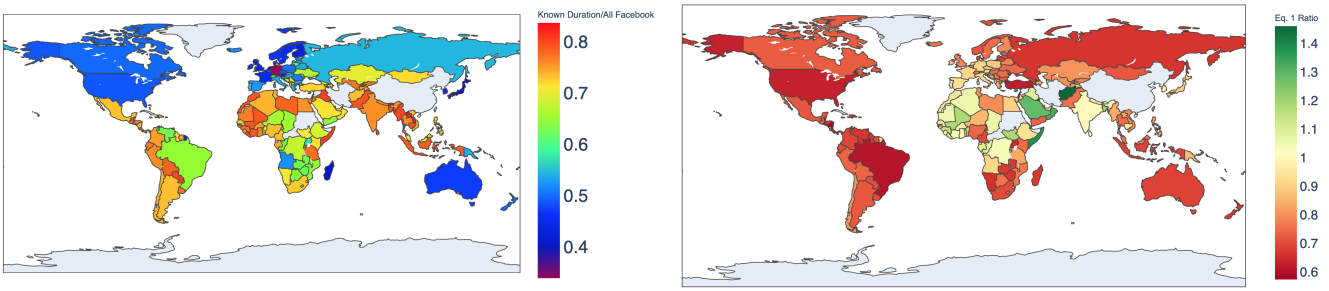
We further use ground truth data [15] on the proportion of individuals with access to mobile phones to show that the proportion of Facebook users matching the "known duration of mobile phone usage" criteria, displays meaningful correlations with mobile phone usage for countries where age-disaggregated estimates for mobile diffusion are available. Comparing changes in Facebook penetration rates (25+ months usage vs. ≤3 months) with ground truth growth (2015 vs. 2018) we find that Facebook data show plausible correlations with changes in smartphone penetration rates for 18-34 year-olds (a correlation (r) of -0.710, n=20 countries). However, the correlations for 50+ year-olds (0.052) indicate a lack of fit for this age group. This could be because "known duration of mobile phone usage" are capturing wider mobile technology diffusion processes, particularly in markets where growth is still occurring.

3.2 Gender Gaps Derived From Known Duration of Mobile Device Usage

We begin exploring gender gaps using this data by looking at cross-county differences in access to mobile phone devices. As we would like to see which countries are showing more growth with respect to gender-equal access to mobile phone devices we calculate the following ratio:

$$\frac{\text{f-to-m ratio for } \leq 3 \text{ months}}{\text{f-to-m ratio for } \geq 25 \text{ months}} \tag{1}$$

The values can be seen in Figure 1b. According to the figure, countries that are currently less gender-equal seem to be comparatively more gender-equal among the most recently joined cohort. In other words, countries that have exceeded gender equality (such as US, Canada, or Switzerland) do not have higher gender ratios among



(a) (b)

Figure 1: Figure 1a: Proportion of Facebook users with known device duration by location. – Figure 1b: Ratio of gender ratios among (i) those who have recently gained access to a mobile device, and (ii) those who have been using a mobile device for over two years. A value < 1.0 (red) indicates that more recent mobile device users on Facebook skew male (compared to long-term users). Similarly, a value > 1.0 (green) indicates that more recent mobile device users on Facebook skew female (compared to long-term users). The ratio displayed in the figure is explained in Equation 1.

more recent joiners compared with long-time users. Whereas countries with considerable digital gender gaps disfavouring women (such as Saudi Arabia or Afghanistan) are now seeing improvements in terms of the gender ratio among recent mobile device users.

We further explore the relationship between the aforementioned metric (Equation 1), with overall gender gaps on Facebook among all users.We find that countries with larger Facebook gender gaps disfavoring women show larger improvements in gender-equality among new joiners with mobile devices, displaying a tendency towards closing of gender gaps. This suggests that in places where gender gaps on Facebook are larger and women are not Facebook users, mobile diffusion processes are helping to close the digital gender gap. Further, the correlation between changes in mobile gender gaps revealed by these data and GDP per capita values (r=-0.17, n=156, p=0.026) and internet penetration (r=-0.4, n=65, p=0.0009) is negative, indicating greater greater equality among new joiners in low-income countries and where internet penetration is lower.

Next we explore one advantage the known duration of mobile device attribute offers, which is to study demographic differences through time using only one snapshot of data – something not possible with the type of targeting attributes used in prior work. More specifically, we present a disaggregation of the female-to-male gender ratios, using the following equation, in terms of both age group and mobile device usage duration.

$$\frac{\#\text{ Female users of age X who have been using a mobile device for Y months}}{\#\text{ Male users of age X who have been using a mobile device for Y months}} \quad (2)$$

Due to space constraints, we only show the obtained patterns for India in Figure 2. But the same analysis can easily be replicated for any country with a sufficient number of Facebook users to allow disaggregation. India has been selected here because it has a large Facebook population with a significant gender gap, but where internet and mobile phone diffusion is still occurring.

Although we can see patterns of gender gaps disfavouring women, Female to Male ratios are often higher among older people than younger people in the country. This is consistent with other work analyzing age and gender differentials on Facebook [9]. This could partially be explained demographically since at older ages there are more women than men in nearly all populations in the world due to the female longevity advantage. On the other hand, these patterns may also reflect differential forms of selectivity and in age-differential patterns of social media use by gender. For example, the sociological literature has argued that women are more likely to be involved in care-giving, providing emotional support, and, in their roles as "kin keepers" in holding family relations together [3, 10]. In a digitalized world where social interaction and communication occurs online, women, including grandmothers, may also be more likely to use social media for kin keeping.

4 DISCUSSION & CONCLUSION

We presented an analysis of a Facebook advertising targeting attribute for users with a known duration of mobile device usage. Though our efforts of understanding what precisely the targeting attribute is measuring could be seen as trying to "reverse engineer" Facebook's black box algorithm, it can also be seen as a case of "algorithmic auditing", i.e. trying to validate that what is claimed is indeed being measured.

Some of our auditing efforts made use of existing sources of "ground truth", e.g. of changes in mobile phone penetration and gender gaps in mobile access. However, even in absence of ground truth with identical definitions, understanding digital behaviors can be valuable in and of itself. Put differently, even if certain patterns in the data apply "only" to the 1.7 billion Facebook users with a known device usage duration, without generalizing to the wider population, then studying these patterns is important for understanding mobile and digital technology diffusion processes in the context of global development. This holds particularly true as this subset of Facebook users seems to skew towards lower income groups and/or countries. For example, we find that these data are better at capturing mobile diffusion processes and primary device use in low-income country contexts. Often in these settings traditional data sources from surveys or censuses on digital access by demographic characteristics such as age and gender may be

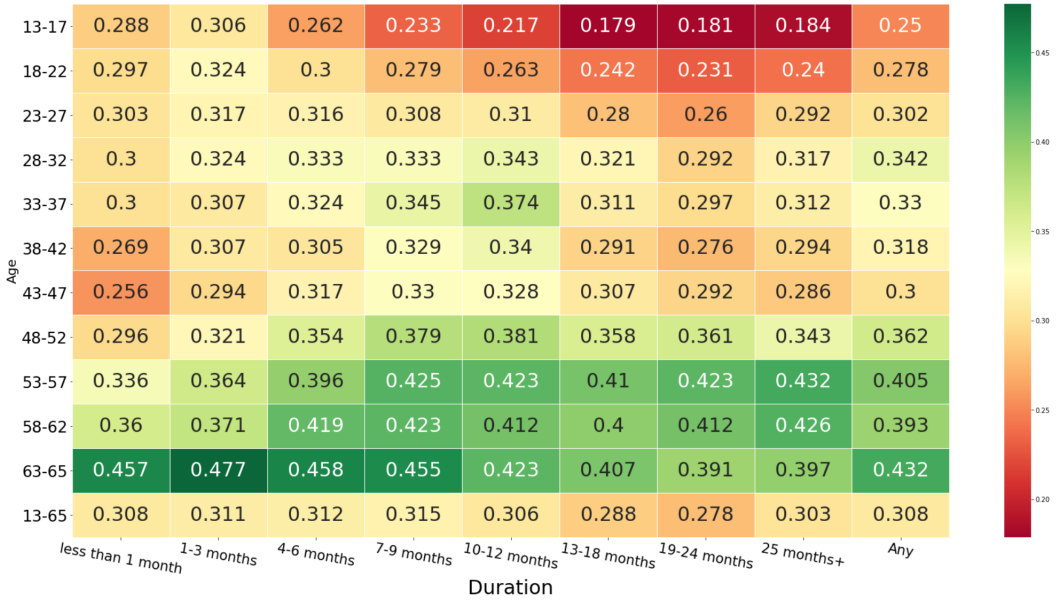


Figure 2: India gender gap with respect to age and duration of using a mobile phone device. Each row represents a duration of access value, and each column represents the age group of the users.

lacking or available only at sporadic durations every few years. These data offer a window into mobile diffusion trends and can help understand short-term dynamics (e.g. over a few months), as well as gender (and age) heterogeneity in these processes.

In future work we would like to explore how the type of data used by us could be applied for nowcasting and for post-filling missing (recent) historic data linked to mobile adoption processes by age and gender. That could be very useful in scenarios where e.g. due to covid-19 data might not be available. At the same time, it is exactly the lack of available historic data to train or calibrate against that makes nowcasting or prediction efforts challenging. Nevertheless, particularly in low-income contexts, these data may reveal interesting insights into changing digital behaviours linked to mobile use for short time-frames than conventional survey estimates are available.

REFERENCES

[1] Monica Alexander, Kivan Polimis, and Emilio Zagheni. 2020. Combining social media and survey data to nowcast migrant stocks in the United States. *Population Research and Policy Review* (2020), 1–28.

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

[3] Piet Bracke, Wendy Christiaens, and Naomi Wauterickx. 2008. The pivotal role of women in informal care. *Journal of Family Issues* 29, 10 (2008), 1348–1378.

[4] Facebook. [n.d.]. Facebook Advertisement Platform. https://www.facebook.com/business/ads.

[5] Masoomali Fatehkia, Benjamin Coles, Ferda Ofli, and Ingmar Weber. 2020. The relative value of facebook advertising data for poverty mapping. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 14. 934–938.

[6] Masoomali Fatehkia, Ridhi Kashyap, and Ingmar Weber. 2018. Using Facebook ad data to track the global digital gender gap. *World Development* 107 (2018), 189–209.

[7] Masoomali Fatehkia, Isabelle Tingzon, Ardie Orden, Stephanie Sy, Vedran Sekara, Manuel Garcia-Herranz, and Ingmar Weber. 2020. Mapping socioeconomic indicators using social media advertising data. *EPJ Data Science* 9, 1 (2020), 22.

[8] David Garcia, Yonas Mitike Kassa, Angel Cuevas, Manuel Cebrian, Esteban Moro, Iyad Rahwan, and Ruben Cuevas. 2018. Analyzing gender inequality through large-scale Facebook advertising data. *Proceedings of the National Academy of Sciences* 115, 27 (2018), 6958–6963.

[9] Sofia Gil-Clavel and Emilio Zagheni. 2019. Demographic differentials in Facebook usage around the world. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 647–650.

[10] Gunhild O Hagestad. 1986. The family: Women and grandparents as kin-keepers. *Our aging society: Paradox and promise* 141 (1986), 60.

[11] Ridhi Kashyap, Masoomali Fatehkia, Reham Al Tamime, and Ingmar Weber. 2020. Monitoring global digital gender inequality using the online populations of Facebook and Google. *Demographic Research* 43 (2020), 779–816.

[12] Yelena Mejova, Harsh Rajiv Gandhi, Tejas Jivanbhai Rafaliya, Mayank Rameshbhai Sitapara, Ridhi Kashyap, and Ingmar Weber. 2018. Measuring subnational digital gender inequality in india through gender gaps in facebook use. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*. 1–5.

[13] Francesco Rampazzo, Emilio Zagheni, Ingmar Weber, Maria Rita Testa, and Francesco Billari. 2018. Mater certa est, pater numquam: what can Facebook advertising data tell us about male fertility rates?. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12. Proceedings of the International AAAI Conference on Web and Social Media.

[14] Filipe N Ribeiro, Fabrício Benevenuto, and Emilio Zagheni. 2020. How Biased is the Population of Facebook Users? Comparing the Demographics of Facebook Users with Census Data to Generate Correction Factors. In *12th ACM Conference on Web Science*. 325–334.

[15] LAURA SILVER. [n.d.]. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally. https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/.

[16] H. Tankovska. 2021. Device usage of Facebook users worldwide as of October 2020. https://www.statista.com/statistics/377808/distribution-of-facebook-users-by-device/.

[17] H. Tankovska. 2021. Global social networks ranked by number of users 2021. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

[18] Ingmar Weber, Ridhi Kashyap, and Emilio Zagheni. 2018. Using advertising audience estimates to improve global development statistics. *Itu Journal: Ict Discoveries* 1, 2 (2018).

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

Visual-Meta Appendix

The data below is what we call Visual-Meta. It is an approach to add information about a document to the document itself, on the same level of the content (in style of BibTeX). It is very important to make clear that Visual-Meta is an approach more than a specific format and that it is based on wrappers. Anyone can make a custom wrapper for custom metadata and append it by specifying what it contains: for example @dublin-core or @rdfs.
The way we have encoded this data, and which we recommend you do for your own documents, is as follows:
When listing the names of the authors, they should be in the format 'last name', a comma, followed by 'first name' then 'middle name' whilst delimiting discrete authors with ('and') between author names, like this: Shakespeare, William and Engelbart, Douglas C.
Dates should be ISO 8601 compliant.
Every citable document will have an ID which we call 'vm-id'. It starts with the date and time the document's metadata/Visual-Meta was 'created' (in UTC), then max first 10 characters of document title.
To parse the Visual-Meta, reader software looks for Visual-Meta in the PDF by scanning the document from the end, for the tag @{visual-meta-end}. If this is found, the software then looks for @{visual-meta-start} and uses the data found between these tags. This was written September 2021. More information is available from https://visual-meta.info for as long as we can maintain the domain.

@{visual-meta-start}

@{visual-meta-header-start}
@visual-meta{version = {1.1},
generator = {ACM Hypertext 21},
organisation = {Association for Computing Machinery}, }

@{visual-meta-header-end}

@{visual-meta-bibtex-self-citation-start}

@article{local-transitory-id,
author = {Sabri, Nazanin and Kashyap, Ridhi and Weber, Ingmar}
title = {Examining Global Mobile Diffusion and Mobile Gender Gaps through Facebook's Advertising Data},
year = {2021},
isbn = {978-1-4503-8551-0},
publisher = {Association for Computing Machinery},
address = {New York, NY, USA},
url = {https://doi.org/10.1145/3465336.3475120},
doi = {10.1145/3465336.3475120},
abstract = {Social media advertising data, particularly data from Facebook's advertising platform, have been successfully used for monitoring population and development indicators, with an emphasis on monitoring digital gender inequality. This paper contributes to this literature by assessing the feasibility of the attribute of user behavior of ``using a mobile device for X months'' available from Facebook's advertising platform to understand short-term global mobile diffusion dynamics and mobile phone gender gaps. We compare this attribute with other features of the platform to form a better understanding of the data and the digital behaviours they capture, and show how this Facebook attribute relates to mobile phone penetration rates and gender gaps in mobile access. We find that this ``Uses a mobile device (X months)'' advertising targeting attribute can be used as a proxy for changes in mobile phone penetration rates, especially among younger users, and that it captures cross-national variation in mobile gender gaps. We further find that countries with larger gender gaps disfavoring women are comparatively more gender-equal among the most recently joined cohort.},
numpages = {4},
keywords = {Gender Gaps; Access to mobile phones; Facebook's Advertisement Platform; Digital Demography; Online Advertising},
location = {Virtual Event, USA},
series = {HT '21}
vm-id = {10.1145/3465336.3475120} }

@{visual-meta-bibtex-self-citation-end}

@{visual-meta-end}