Leveraging online advertising data for measuring the sustainable development goals: applications for gender gaps and SDG 5

Reham Al Tamime, Masoomali Fatehkia, and Ingmar Weber (Qatar Computing Research Institute, Hamad Bin Khalifa University); Ridhi Kashyap (University of Oxford)

Abstract

In 2015, the United Nations Member States adopted the 2030 Agenda for Sustainable Development. At the heart of this are the 17 Sustainable Development Goals (SDGs), which call for action by all countries through a global partnership. The need to better monitor progress towards the realization of the SDGs has led to a concurrent call for a data revolution, including the need to harness non-traditional, 'big data' sources. The need to consider new data streams, emerging from new technologies, is amplified by the fact that traditional data sources like surveys and censuses are resource-intensive, and lack geographical and temporal resolution, especially in low- and middle-income countries. Here we describe how a novel stream of 'big data' – a new data source: aggregate, anonymous online advertising audience estimates provided by companies such as Facebook, Google, LinkedIn, and others – can help fill existing data gaps and provide real-time estimates of SDG-relevant indicators and processes. We provide examples focusing on SDG 5, which seeks to promote gender equality and promote the empowerment of women and girls.

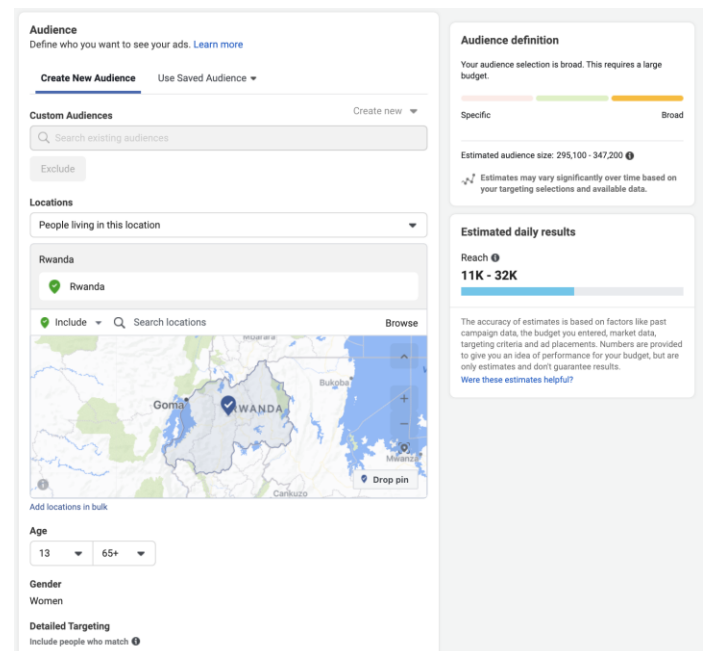## ONLINE ADVERTISING AUDIENCE ESTIMATES

Online advertising is the main revenue source for large online platforms and social media companies. During the process of setting up an advertising campaign on their platform, Facebook, Google, and LinkedIn provide advertisers with services to choose their audience, and manage ad budgets. These services also allow advertisers to estimate the potential audience reach for a given set of targeting criteria. *Figure 1. Facebook Ads Manager* shows an example screenshot of Facebook's Ads Manager[1]. The screenshot shows on the right the "Potential reach", 295,100 (min) - 347,200

(max) people, when targeting female Facebook users located in Rwanda, aged 13 to 65+. Similar interfaces are provided by the advertising platforms of Google, LinkedIn and Snapchat, though the supported targeting criteria differ by platform. Most platforms support retrieving audience reach estimates programmatically by providing an application programming interface (API).

These estimates are available in near real-time and can be filtered by age, gender, location, as well as other characteristics. In order to prevent identifications of users, online platforms do not report the exact number of users matching the targeting criteria when the matching audience size is below a threshold. Some platforms, in particular Facebook, report estimates for both daily active users (DAUs) and monthly active users (MAUs). DAUs are defined as the "estimated number of people that have been active on your selected platforms

and satisfy your targeting spec in the past day"[2], while MAUs are defined as the "estimated number of people that have been active on your selected platforms and satisfy your targeting spec in the past month"[3]. In the following, we describe two examples that use online advertising audience estimates from Facebook, Google, and LinkedIn.

Figure 1. Facebook Ads Manager

[1] https://www.facebook.com/business/tools/ads-manager

[2] https://developers.facebook.com/docs/marketing-api/reference/ad-campaign-delivery-estimate/

CASE STUDY1: MODELLING SDG5 RELEVANT INDICATORS USING ONLINE ADVERTISEMENT DATA

One of the targets in SDG 5 is to "enhance the use of enabling technology, in particular information and communications technology (ICTs), to promote the empowerment of women"[3]. Even though the use of the Internet and mobile phones has rapidly increased, a digital gender divide persists across many countries where women access the Internet and own mobile phones less than men, both reflecting and exacerbating other gender inequalities. According to the UN's specialized agency for ICTs, the International Telecommunication Union (ITU), over 250 million fewer women are online than men, and gender gaps in internet use tend to be greater in low- and middle-income countries [1]. In order to overcome the digital gender divide, several organisations including ITU, GSMA, and UNESCO have developed a set of recommended actions designed to address the gender gap in Internet and broadband access and use. However, many countries do not conduct any surveys on Internet access and use, which makes it hard to measure progress on digital gender equality, and makes it hard to formulate policies to improve it.

To help fill these data gaps in measuring global gender disparities in internet use and mobile access, previous work used Facebook advertising data [2]. In particular, audience reach estimates obtained through the Facebook advertising platform were disaggregated by age and gender to predict digital gender gaps for more than 150 countries. Building on the aforementioned study, other work has used an additional data source, Google's advertisement impression estimates in combination with Facebook's, to further improve coverage of global digital gender gaps [3]. These studies find that Facebook and Google Ads data are both highly correlated with available official survey-based statistics on internet and mobile gender gaps (e.g. from the ITU). In addition, Facebook and Google Ada data are correlated with the ITU data on low-level digital skills such as using copy and paste tools, transferring files, and sending emails. *Figure 2* shows global gender gaps in internet access. Figure 2a is the Internet Gender Gap Index data computed using the ITU data, while Figure 2b shows the ratio of female-to-male internet use predicted using the country-specific Facebook Gender Gap Index. Facebook Ads data can be used to predict internet gender gaps for a large number of countries,

especially in South Asia and Africa for which no gender gap data from ITU are available.

Figure 2. The internet gender gap index (a) computed using ITU ground truth data, and (b) predicted using Facebook Ads data. Interactive version available at: https://www.digitalgendergaps.org
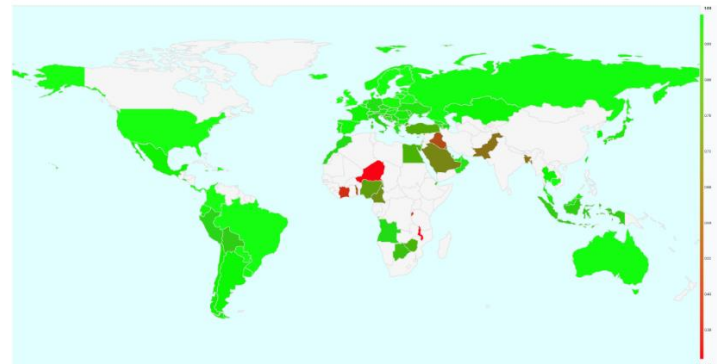
(a) Internet Gender Gap Index Computed using the ITU Data
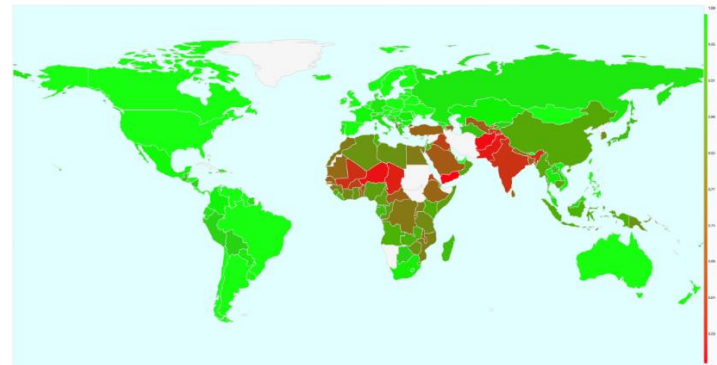
(b) Internet Gender Gap Index Predicted using the Facebook data

In related work, the Facebook gender gap, generated using data available from Facebook Ads, has been shown to be correlated with gender gaps in education and economic opportunity. This suggests that online gender gaps may be connected with broader offline gender inequalities [4]. In applications to other domains, work using LinkedIn Ads data has examined gender gender inequalities in the professional domain [5]. Specifically, LinkedIn Ads data was disaggregated by age, gender, company, industry, job seniority, job function, field of study, and country in order to model global gender gap indicators linked to women's economic empowerment. Similar data was used to study gender skill gaps in the US context [6].

CASE STUDY2: MONITORING REAL-TIME CHANGES

The examples presented so far focus on relatively slower moving gender trends. However, online advertising data can also be used in more fast-moving humanitarian situations where real-time data are

---

[3]https://stats.unctad.org/Dgff2016/people/goal5/target_5_b.html

useful, but where standard modes of data collection are unavailable. As a proof of concept, we demonstrate trends of Facebook Daily Active Users (DAU) in Afghanistan which in August 2021 experienced a collapse of its Western-backed government following the fall of Kabul to the Taliban forces. How did the return of the Taliban government, known for its hostility to women's social, economic and political participation, impact on women's participation online? Did a potential increase in offline restrictions arising from political events result in a further retreat for women online, or did social media provide a platform for a new type of social mobilization and expression?

Figure 3 shows trends for adolescent girls aged 15-19 on Facebook as well as the female-to-male ratio of users in that age group. We choose to illustrate this age group as this generation never experienced Taliban rule during their lifetime, whilst having come of age at a time when social media and digital technologies have become more widely available. As can be seen in the figure, the beginning of the Taliban offensive and its aftermath witnessed increases in social media activity on Facebook among adolescent girls to a level not seen in the recent past. However, more recently this increased usage level seems to be declining. Based on this data, there appears to have been an overall more positive change in the presence of girls online in the short term, indicative of a form of digital mobilization or protest response; continuous monitoring of these trends is required in order to study more long-term changes.

While we do not have subnational historical data, based on recent estimates (as of 20 Jan 2022) about two-thirds of the girls aged 15-19 are from Kabul (21,870 DAU in Kabul out of 34,042 in the entire country). Hence, it is important when interpreting the country-level data to keep in mind the bias towards more urban areas of the country which may not reflect the situation in rural areas. Going forward, we are also collecting subnationally disaggregated data for Afghanistan.

Figure 3. Trends for teenage girls aged 15-19 as well as the female-to-male ratio of users in that age group. Timeline of selected dates for context. A. 28 Sep 2019, Afghanistan 2019 presidential elections. B. 29 Feb 2020, Doha Agreement signed between the Taliban and the USA. C. 1 May 2021, beginning of 2021 Taliban offensive. D. 15 Aug 2021, fall of Kabul to the Taliban. E. 12 Sep 2021, Afghan women engage in #DoNotTouchMyClothes campaign as social media protest.
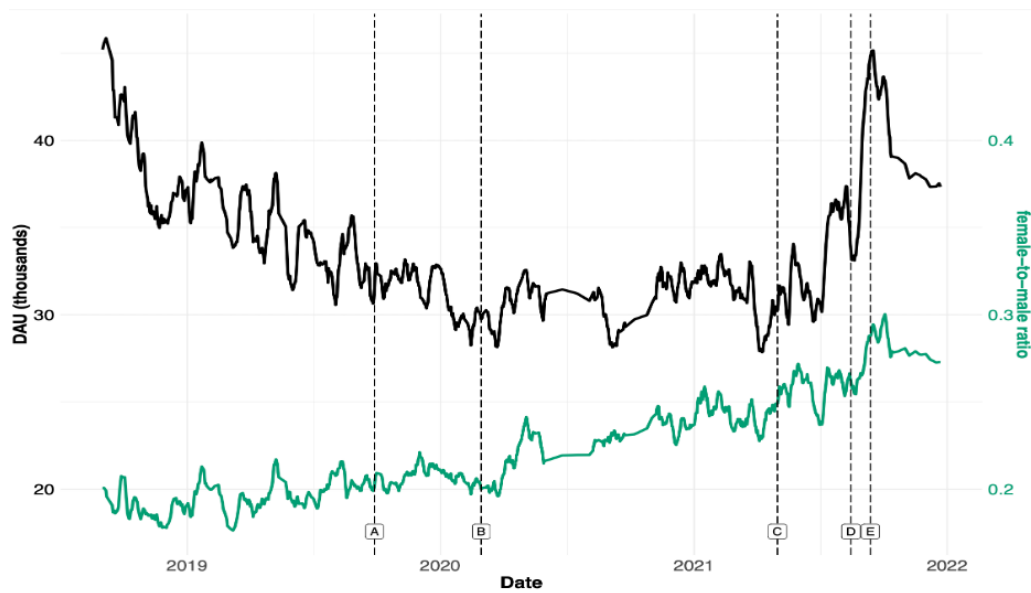
LIMITATION AND CHALLENGES

The above case studies demonstrated the value of online advertising data for monitoring global sustainable development, particularly linked to SDG 5. However, there are limitations and challenges that are important to highlight with this approach. First, social media companies lack transparency about how these advertising data are generated. In particular, companies like Facebook do not use a single algorithm, but rather "multiple layers of machine learning models and rankings"[4], in particular for inferring certain attributes (e.g. related to interests). These algorithms work as a

---

[4]https://economictimes.indiatimes.com/tech/trendspotting/explained-the-algorithms-that-run-facebook/articleshow/87282136.cms

black box without an easy way for academics or policy makers to audit aspects related to data quality. In addition, these algorithms are constantly changing, as they are designed to meet advertising and not policy or research needs, which can make it hard to obtain consistent and comparable signals over time. For example, Palotti et al. observed a sudden change in March 2019 in how Facebook classifies users by countries they have lived in [7]. Finally, the issue of user awareness of privacy issues is important to consider. Specifically, social media companies shadow users as they like pages, comment, or browse websites. As companies profile users, people might not realize that their aggregate data can be accessed and used for advertisement or academic research. Greater public awareness of how data are collected by companies, as well as the development of transparent and privacy-preserving models for data sharing between companies, researchers and policy-makers, are desirable for using these technologies in an ethical and sustainable manner.

References

[1] International Telecommunication Union, "Fast-forward progress leveraging tech to achieve the global goals," 2017. [Online]. Available: https://www.itu.int/en/sustainable-world/Documents/Fast-forward_progress_report_414709%20FINAL.pdf

[2] M. Fatehkia, R. Kashyap, and I. Weber, "Using Facebook ad data to track the global digital gender gap," *World Dev.*, vol. 107, pp. 189–209, Jul. 2018, doi: 10.1016/j.worlddev.2018.03.007.

[3] R. Kashyap, M. Fatehkia, R. Al Tamime, and I. Weber, "Monitoring global digital gender inequality using the online populations of Facebook and Google," *Demogr. Res.*, vol. 43, no. 27, pp. 779–816, Sep. 2020, doi: 10.4054/DemRes.2020.43.27.

[4] D. Garcia *et al.*, "Analyzing gender inequality through large-scale Facebook advertising data," *Proc. Natl. Acad. Sci.*, vol. 115, no. 27, pp. 6958–6963, Jul. 2018, doi: 10.1073/pnas.1717781115.

[5] R. Kashyap and F. C. J. Verkroost, "Analysing global professional gender gaps using LinkedIn advertising data," *EPJ Data Sci.*, vol. 10, no. 1, p. 39, Jul. 2021, doi: 10.1140/epjds/s13688-021-00294-7.

[6] K. Haranko, K. Garimella, E. Zagheni, and I. Weber, "Professional gender gaps across US cities," *Proc. of the Intern. Conf. on Web and Social Media*, pp. 604-607, 2018.

[7] J. Palotti, N. Adler, A. Morales-Guzman, J.Villaveces, V. Sekara, M. Garcia Herranz, M. Al-Asad, and I. Weber, "Monitoring of the Venezuelan Exodus through Facebook's advertising platform," *PLOS ONE*, 15(2): e0229175.