Using Facebook ad data to track the global digital gender gap

Masoomali Fatehkia [a], Ridhi Kashyap [b,*], Ingmar Weber [c]

[a] *Princeton University, USA*
[b] *University of Oxford, UK*
[c] *Qatar Computing Research Institute, Qatar*

ARTICLE INFO

ABSTRACT

Gender equality in access to the internet and mobile phones has become increasingly recognised as a development goal. Monitoring progress towards this goal however is challenging due to the limited availability of gender-disaggregated data, particularly in low-income countries. In this data sparse context, we examine the potential of a source of digital trace 'big data' – Facebook's advertisement audience estimates – that provides aggregate data on Facebook users by demographic characteristics covering the platform's over 2 billion users to measure and 'nowcast' digital gender gaps. We generate a unique country-level dataset combining 'online' indicators of Facebook users by gender, age and device type, 'offline' indicators related to a country's overall development and gender gaps, and official data on gender gaps in internet and mobile access where available. Using this dataset, we predict internet and mobile phone gender gaps from official data using online indicators, as well as online and offline indicators. We find that the online Facebook gender gap indicators are highly correlated with official statistics on internet and mobile phone gender gaps. For internet gender gaps, models using Facebook data do better than those using offline indicators alone. Models combining online and offline variables however have the highest predictive power. Our approach demonstrates the feasibility of using Facebook data for real-time tracking of digital gender gaps. It enables us to improve geographical coverage for an important development indicator, with the biggest gains made for low-income countries for which existing data are most limited.

1. Introduction

The rapid proliferation of information and communication technologies (ICTs) has been one of the most significant social phenomena of the new millennium. Today, there are estimated to be more than 7 billion mobile subscriptions worldwide, up from 738 million in 2000. Globally, the number of internet users has risen from roughly 400 million in 2000 to 3.2 billion people who are using the Internet today. 2 billion of these users live in developing countries (ITU, 2015). The expansion of ICTs has paved the way for the digital revolution that is shaping the ways in which people and communities learn and access information and skills, connect with each other, express their ideas, and conduct their everyday lives.

The tremendous potential of digital technologies as tools for empowering communities in poor countries and delivering development goals has become widely recognized among development practitioners (Qiang, Clarke, & Halewood, 2006; Unwin, 2009; Walsham & Sahay, 2006). By lowering the costs of information and connectivity, digital technologies can improve employment opportunities, provide cost-effective health services, and enable access to learning, skills and financial services to help achieve sustainable development goals, particularly for marginalized groups such as women (Broadband Commission, 2013; Hafkin & Huyer, 2006; Huyer & Carr, 2002; Santosham & Lindsey, 2015; WWW Foundation, 2015). This commitment is explicitly noted in the United Nations (UN) Sustainable Development Goals (SDGs), in which Goal 5(b) pledges to "enhance the use of ...information and communication technology to promote the empowerment of women".[1]

Although significant reductions in the costs of ICTs have facilitated their rapid expansion in the developing world in the past two decades, significant gender inequalities in access to these technologies persist. The International Telecommunications Union (ITU), the UN's specialized agency for ICTs, estimates that some 200 million fewer women are online compared with men (Broadband Commission, 2013). Although the digital divide, a term used to describe inequalities in access to the internet, between

---

[1] https://sustainabledevelopment.un.org/post2015/transformingourworld.

men and men is not restricted to developing countries, gender inequalities in access are greater in the developing world (Alozie & Akpan-Obong, 2017; Broadband Commission, 2013; Hafkin & Huyer, 2007). As a result, concerns about gender inequalities in internet and mobile access have emerged as an important focus of the discussions on ICTs for development (Broadband Commission, 2013; Moolman, Primo, & Shackleton, 2007; Santosham & Lindsey, 2015). In March 2013, the International Telecommunications Union (ITU) and UNESCO Broadband Commission for Digital Development further endorsed a target calling for gender equality in access to broadband by 2020 (Broadband Commission, 2013).

Despite the increasing visibility of the issue, our ability to measure gender gaps in access to the internet and mobile phones is significantly limited due to data gaps. A report by the ITU/UNESCO identified the lack of sex-disaggregated data on internet and mobile phone access as "one of the key barriers" for measuring progress in development goals that call for gender equality in access to the internet (Broadband Commission, 2013, p. 19). Official, nationally representative gender-disaggregated statistics on internet and mobile access lack coverage across countries and regular production, and data availability on these indicators is especially limited in low-income countries (Antonio & Tuffley, 2014; Brännström, 2012; Hafkin & Huyer, 2007). When data are available, inconsistencies across data sources and measures impede cross-national comparisons. Furthermore, these data lack granularities at the subnational level or by socio-demographic group, such as age or education.

In this data sparse context, there is the need to look at other data sources that can be used to measure and track the current state of digital gender disparities. The continually updating nature of 'big data' sources, particularly digital trace data, that capture online footprints left behind on digital spaces in real time offer promise in this setting for nowcasting, or in other words generating real-time predictions of social outcome indicators in the present (di Bella, Leporatti, & Maggino, 2018). Nowcasting is typically employed when the actual value of indicator of interest will only be known with a significant delay, creating the need to "predict the present" (Blumenstock, Cadamuro, & On, 2015; Mao, Shuai, Ahn, & Bollen, 2015; Elvidge et al., 2009; Evangelos, Efthimios, & Konstantinos, 2013; Giannone, Reichlin, & Small, 2008; Harald et al., 2013; Lampos & Cristianini, 2012; Yazdani & Manovich, 2015).

This paper leverages one such source of digital trace data – Facebook's advertisement audience estimates – to measure and nowcast digital gender gaps in internet and mobile access in a global perspective. We generate a unique country-level dataset combining 'online' indicators of Facebook users by gender, age and device type, 'offline' indicators related to a country's overall development and gender-specific development indicators, and official data on gender gaps in internet and mobile access where available. As not all those with access to the internet and mobile phones are Facebook users, we need to examine the validity of Facebook-derived data for measuring digital gender gaps more broadly. To do this, we predict gender gaps in internet and mobile phone access from official data using online, Facebook-derived indicators, as well as models combining online and offline indicators.

Our results demonstrate the feasibility of using Facebook data to measure digital gender gaps around the world. Models using only online Facebook data do better than models using only offline indicators for predicting internet gender gaps as reported in official ITU statistics. Models combining Facebook data with offline gender equality indicators however do the best. With the online model using Facebook data, we estimate a mean value of the internet gender gap index for low-income countries to be 0.76. This implies that 0.76 women have access to the internet for every man who

does, indicating that female internet penetration is 24% lower than that for males. This index increases to 0.86 for lower-middle income countries and to 0.92 and 0.95 for upper-middle and high income countries respectively based on the World Bank's classification. For the mobile phone gender gaps our online model predicts a gender gap index of 0.79 for low income countries indicating a female deficit in mobile access of 21%. This increases to 0.90 and 0.96 for lower-middle and upper-middle income countries respectively, with high income countries having an index of 0.98.

The Facebook data enable us to improve coverage for indicators of the internet gender gap significantly from 84 countries,[2] for which official data are available from the ITU, to 152. The biggest gains are for low- and lower-middle-income countries, for which the Facebook data enable us to generate measures for 64 countries compared with 13 in the ITU data. For mobile phone gender gaps, the Facebook data enable us to expand coverage from 22 countries to 153, with the biggest gains again for low- and lower-middle-income countries where we are able to generate measures for 64 countries instead of the 17 in currently available sources. Our work presents an example of data innovation that harnesses the potential of 'big data' sources for real-time and global monitoring of development indicators (IEAG, 2014; Letouze & Jutting, 2014).

2. Digital gender gaps

2.1. The data gap: gender and ICTs

The 'digital divide' is a concept that has been used to describe inequalities in access to ICTs, in particular access to and use of the internet (Norris, 2001). These could refer to disparities across different dimensions, for example between developed and developing countries, as well as those between groups within countries, such as those between men and women, or the rich and the poor. In this paper, our focus is on the digital gender gaps, specifically gaps in access to the internet and mobile phones between men and women. Digital gender gaps can be of various forms, including differences in access, the extent of use, in technical skills, and social support in using these technologies (Broadband Commission, 2013; Bimber, 2000). Given significant variation in internet and mobile penetration rates, global analyses of the digital gender gap have predominantly referred to the gaps in access and use of the internet without reference to specific types of uses or skills.

As with any technology, the expansion of ICTs has followed a diffusion process with early adopters and late adopters (Rogers, 2010). Studies from industrialized countries such as the US and Canada have shown that patterns of early adoption online often mirrored social inequalities from the offline world, as high income, educated and urban groups, and men were more likely to go online earlier (DiMaggio, Hargittai, Neuman, & Robinson, 2001). The increasing diffusion of digital technologies in industrialized countries has generally accompanied a reduction in gender gaps in their access (Haight, Quan-Haase, & Corbett, 2014; Ono & Zavodny, 2007; Rice & Katz, 2003).

Although gaps in access have closed, differentiated patterns of use between men and women have emerged (DiMaggio, Hargittai, Celeste, & Shafer, 2004). Men remain more frequent users of the internet, reporting higher levels of daily usage, and are more likely to use the internet for activities such as games, music and online trading, whereas women were more likely to use the internet for instant messaging and staying connected

---

[2] Of these 84, we obtained Facebook ad audience data for 78 countries. Three countries, Iran, Cuba and Sudan, are not currently supported by Facebook. Three other countries, Montserrat, Palestine and Puerto Rico, were not supported in April 2017 when we compiled the list of countries to consider.