

# Tracking global gender gaps in information technology using online data

*By Florianne C.J. Verkroost, Ridhi Kashyap, Kiran Garimella, Ingmar Weber and Emilio Zagheni*

## Introduction

Despite the proliferation of information and communication technologies (ICTs), significant gender inequalities in the use of these technologies persist (Fatehkia et al., 2018). The International Telecommunication Union (ITU) (2017) estimates that around 200 million fewer women are online than men. While gender gaps disfavouring women in Internet access and basic digital skills can be sizeable, especially in less developed countries, gender gaps in high-level digital skills, such as those required to work in jobs in the information technology (IT) sector, are thought to be even larger in most countries (Borgonovi et al., 2018). For example, a recent report showed that only 16 per cent of the talent in artificial intelligence, an area of significant projected growth in the IT sector, is female (Roca, 2019). These gender gaps are especially important to measure and understand in the context of increasing digitization of labour markets, and to ensure the continued economic empowerment of women as envisaged in Sustainable Development Goal 5 on gender equality (Davaki, 2018). Furthermore, with the increasing diversification and specialization of jobs in the IT sector, it is important to understand the variation in gender gaps across different subdomains within the sector.

This article contributes to understanding global variations in gender gaps in IT industries by drawing on a novel source of online data – specifically, aggregate, anonymous data from the advertising platform of LinkedIn, the world’s largest professional networking website. LinkedIn provides an advertising platform called LinkedIn Campaign Manager, which can be used to create and manage advertisements. Potential advertisers can specify their desired audience by providing targeting criteria such as gender, geography (e.g. country, city) and job industry. Based on these features, and before an advertisement is launched,

LinkedIn provides an estimate of how many female and male LinkedIn users work in a particular industry per country, giving us an estimate of the supply of these industries.<sup>1</sup> By leveraging these aggregate data on the number of LinkedIn users, we examine the variation across countries in supply-side gender gaps in different subdomains of the IT sector (e.g. computer hardware or software, computer and network security, etc. as defined by LinkedIn Campaign Manager).

## Data and methods

### LinkedIn’s advertising platform

LinkedIn is the world’s biggest professional social networking website and currently has about 660 million users.<sup>2</sup> LinkedIn offers targeted advertising on the platform, whereby audiences can be targeted on demographic criteria (e.g. geography, gender, age group), as well as professional criteria (e.g. job seniority, company industry). Advertisers are shown approximations of the targeted audience size before launching advertisements. This aggregate count of users meeting a particular set of criteria essentially functions as a ‘digital census’ from a social research perspective, which enables us to collect data from a sizeable population that are available in real-time and capture about 17 per cent of all roughly 4 billion Internet users (International Telecommunication Union, 2019). These data have been used previously, for example to study labour migration patterns into the United States (State et al., 2014) and gender gaps in the US labour market (Harankó et al., 2018).

Although LinkedIn is a platform for skilled professionals, which makes it well-suited to study professional gender gaps, it is also an online population whose characteristics may differ from the actual labour force. For example, the

LinkedIn population may be younger than the actual labour force, because LinkedIn is an online platform that is likely used more by younger and/or more Internet-connected individuals. If there are systematic gender inequalities in Internet usage in a country, gender gaps on LinkedIn may reflect gender gaps in Internet use/access rather than professional gender gaps. We expect this bias to be relatively weak, however, given that we are interested in a specialized population of those employed in IT. Moreover, to address some of the bias associated with the fact that LinkedIn does not capture the actual labour force, we focus our analysis on relative female-to-male ratios rather than absolute user counts. LinkedIn data also have some advantages over labour-force survey data: LinkedIn data offer a global, comparable and harmonized perspective; have greater country coverage; have the potential to be more up-to-date than labour survey data; and can be collected frequently. Furthermore, the data can be disaggregated on the basis of many demographic and professional characteristics, allowing us to explore and explain gender gaps in different subgroups of the LinkedIn population.

### Data and analysis outline

The data collected from LinkedIn's advertising platform contain about 371 million users for which country, gender and industry are available. Of these users, almost 43 million (11 per cent) report that they are employed in an IT-related industry (hardware, computer networking, semiconductors, telecommunications, wireless, computer software, Internet-related industries, IT and services, and computer and network security). LinkedIn users are classified into such industries based on the primary industry in which the company they claim to work for on their profile operates.<sup>3</sup> LinkedIn uses its own industry taxonomy to classify company industries (Fang, 2016), which is roughly equivalent to International Standard Industrial Classification (ISIC), albeit with some exceptions, such as the Internet industry (Zhu et al., 2018). Data on audience estimates by gender and industry are available for 181 countries. Using these aggregate counts, we analyse global variations in gender gap indices (GGIs) for different subsets of the LinkedIn population, such as those in IT industries, specific subdomains of the IT industry, and the overall (IT and non-IT) LinkedIn population. A GGI value of one would indicate parity between men

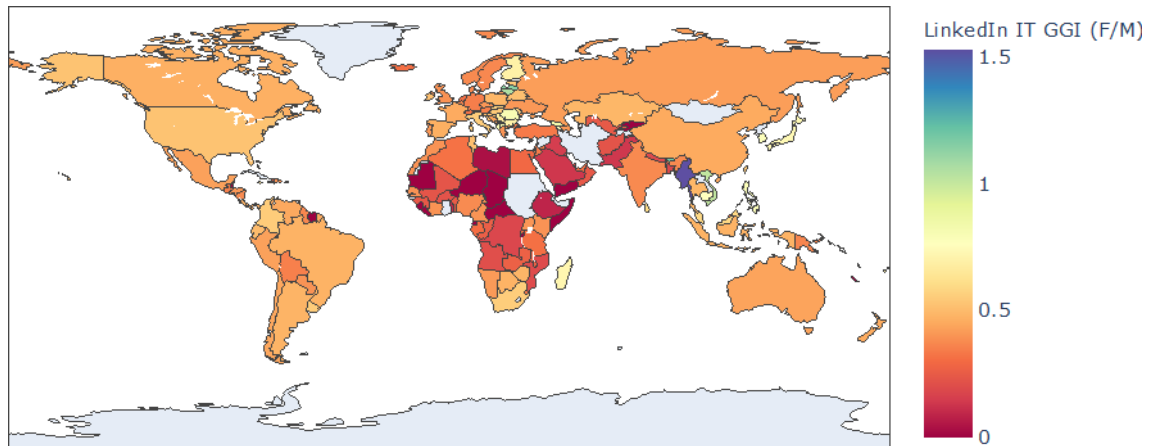
and women; a value less than one indicates that women are underrepresented relative to men; and a value greater than one indicates that women are overrepresented relative to men.

We further compare LinkedIn IT GGIs with those on other platforms such as Facebook. Previous research has shown that gender gaps in Facebook use are a good proxy of country-level Internet use gender gaps (Fatehkia et al., 2018). We also examine to what extent LinkedIn IT GGIs can be explained by development indicators such as gross domestic product (GDP) per capita (World Bank Group, 2020), as well as gender-specific development indicators such as composite indicators of gender gaps in educational attainment and economic opportunity (World Economic Forum, 2019), and other gender gap indicators in education and employment, which we compute ourselves. These include gender gaps in education in both the science, technology, engineering and mathematics (STEM) and ICT domains (United Nations Educational, Scientific and Cultural Organization (UNESCO), 2019), in technical and professional employment (International Labour Organization (ILO), 2019b),<sup>4</sup> and in labour force participation rates (ILO, 2019c). We finally compare the LinkedIn IT GGI to the ILO IT GGI, which is a gender gap indicator from ILO computed from the female-to-male ratio in employment in ISIC divisions 58 to 63 of section J, 'Information and communication' (ILO, 2019a; United Nations Department of Economic and Social Affairs, 2008).

### Gender gaps in IT industries

What do LinkedIn gender gaps in IT look like across the globe? Figure 8.1 shows the geographic distribution of the LinkedIn GGI in IT industries. For the countries coloured light blue/grey, LinkedIn audience estimates are not available. Additionally, LinkedIn has been blocked in some countries (e.g. China, the Russian Federation and the Islamic Republic of Iran). Therefore, estimates for these countries may be biased (as discussed later). From this map, we firstly observe that in the large majority of countries worldwide, there are more men than women in IT, as indicated by GGIs smaller than one. While there are generally twice as many men as women (GGI = 0.5) in most of the Americas, Europe, Asia and Oceania, there is even more gender inequality in IT industries in

Figure 8.1. World map of the GGI (female/male) of LinkedIn users in IT industries



Source: Authors own 2020, unpublished, using data from LinkedIn Campaign Manager<sup>5</sup>

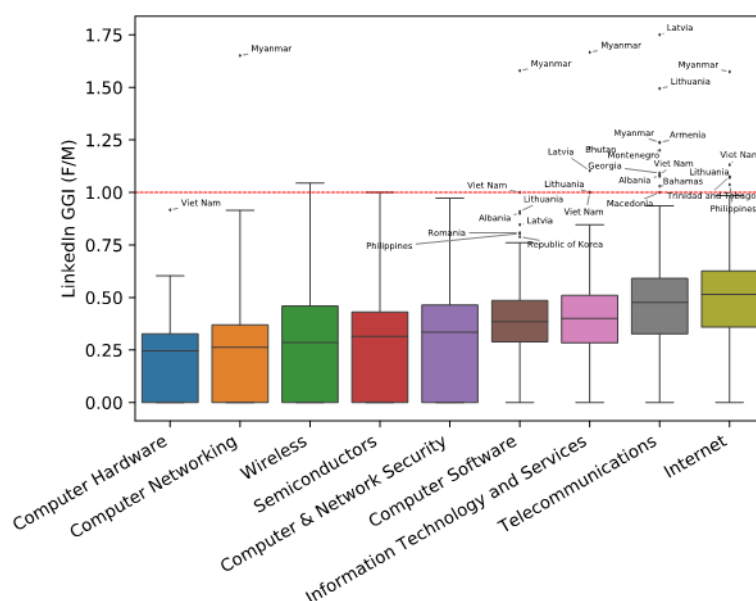
Africa with GGIs approaching zero, especially in the northern and central regions of Africa. There are a few exceptions: in Myanmar (1.53), Bhutan (1.21), Latvia (1.15), Lithuania (1.06) and Viet Nam (1.02), there are more women than men in IT industries on LinkedIn.

Are there any differences in gender inequality across various IT industries? Figure 8.2 shows a box and whisker plot for each of the IT industries in the LinkedIn data. The box ranges between the first and third quartile; the horizontal line in each box represents the median; the whiskers represent the minimum and maximum; and the

sole dots represent outliers. The red horizontal line at GGI equals one shows the 'equality cut-off', and we observe that all IT industries have gender ratios far below this cut-off. While there are about four times more men than women (median GGI = 0.25) in computer hardware and computer networking, there are about twice as many men as women (median GGI = 0.5) in Internet and telecommunications.

It seems that generally there is more gender inequality in the industry defined by the United Nations Department of Economic and Social Affairs (2008) as 'computer programming, consultancy

Figure 8.2. Box and whisker plot of the GGI (female/male) of LinkedIn users in IT industries across countries

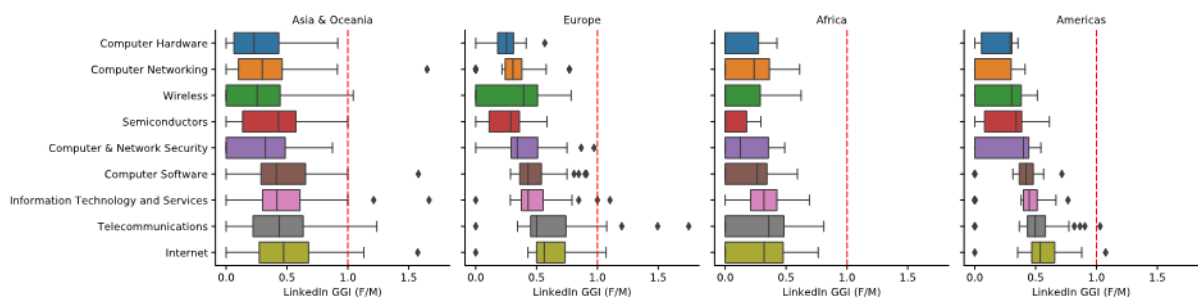


Source: Authors own 2020, unpublished, using data from LinkedIn Campaign Manager<sup>5</sup>

and related activities', than in industries classified as 'telecommunications' and 'information service activities'.<sup>5</sup> This may well be because, for the two sectors, the former is more related to computational work and programming, a sub-area where women's underrepresentation has been noted (Lipowiecka & Kiriti-Nganga, 2016), while the latter is focused more on communications and services, an area within ICT where women tend to be better represented (Lipowiecka & Kiriti-Nganga, 2016). Again, there are some exceptional countries for which gender equality is (almost) achieved in IT, as shown per industry in Figure 8.2. Some countries recur across these industries, such as Albania, Latvia, Lithuania, Myanmar and Viet Nam. Additionally, the telecommunications and computer software industries tend to be more gender egalitarian in, respectively, Armenia, Montenegro, Georgia, the Philippines, Romania and the Republic of Korea.

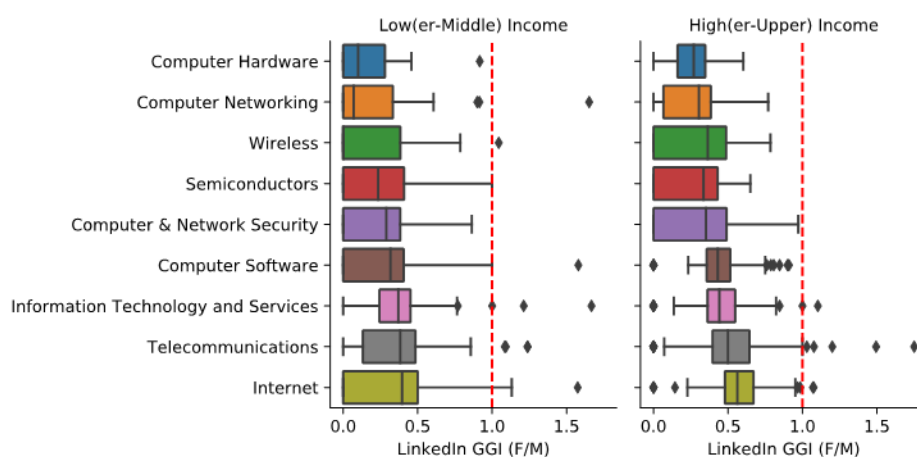
Figures 8.3 and 8.4 show the same box and whisker plot as in Figure 8.2, now faceted by geographic and income regions, respectively. The geographic regions are based on the M49 standard of the United Nations<sup>7</sup> and have been merged into more aggregate regions with similar sample sizes. These figures show that the highest gender inequality in IT on LinkedIn occurs in Africa, while Europe and the Americas have the smallest variation in IT gender gaps across the countries in their regions. Additionally, there are some countries in Asia, Oceania, Europe and the Americas where there are more women than men in IT, while all countries in Africa have more men than women in IT. Furthermore, high and upper-middle income countries have more gender egalitarian IT sectors than low and low-middle income countries. The patterns in gender gaps among the IT industries are similar across geographic and income regions – for example, median gender gaps are larger in IT

**Figure 8.3. Box and whisker plot of the GGI (female/male) of LinkedIn users in IT industries across countries, faceted by geographic region**



Source: Authors own 2020, unpublished, using data from LinkedIn Campaign Manager<sup>6</sup> and United Nations Department of Economic and Social Affairs Statistics Division<sup>6</sup>

**Figure 8.4. Box and whisker plot of the GGI (female/male) of LinkedIn users in IT industries across countries, faceted by income region**



Source: Authors own 2020, unpublished, using data from LinkedIn Campaign Manager<sup>6</sup> and World Bank Group<sup>6</sup>

industries such as computer and network security and computer hardware, and they are smaller in the Internet and IT and services industries.

## Women in IT industries: Exploring selection effects

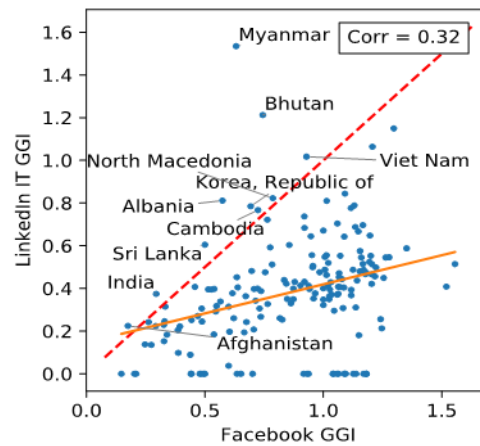
We now turn to compare LinkedIn gender gaps with other gender gap indicators. Figure 8.5 firstly shows a comparison between the LinkedIn IT GGI and the Facebook GGI of users aged 18+ (FB GGI), an indicator of the female-to-male ratio of monthly active Facebook users, which Fatehkia et al. (2018) have shown to be strongly predictive of ITU Internet access gender gaps. Another advantage of the FB GGI is that it provides better geographic coverage than ITU data.<sup>8</sup> Second, we compare the LinkedIn IT GGI with the overall LinkedIn GGI aggregated across all industries on the platform (Figure 8.6). These comparisons help to address two distinct questions. First, how do high-level digital gender gaps such as working in an IT industry, as approximated by the LinkedIn IT GGI, compare with Internet access or low-level digital gender gaps, as approximated by the FB GGI? Second, how do gender gaps in IT on LinkedIn compare with other industries more generally, as captured by the overall LinkedIn GGI?

From Figures 8.5 and 8.6, we observe that the magnitude of gender inequality in LinkedIn IT industries is larger than both gender inequality on Facebook and gender inequality on LinkedIn aggregated across all industries for most of the countries (points lie below the red dashed  $x=y$  line). However, for a handful of countries in South Asia (e.g. India, Myanmar), gender equality in IT is greater than in all industries combined on LinkedIn. In the comparison with Facebook gender gaps, these aforementioned countries – as well as Albania, Bhutan, Cambodia, North Macedonia, Sri Lanka and Viet Nam – emerge

as those where there is greater gender equality in LinkedIn IT industries than in the Facebook population. In these countries, we observe a positive selection effect in the IT sector, with more women represented than would be predicted by the Internet or digital access gender gaps (proxied from Facebook usage gaps) and professional gender gaps (proxied from LinkedIn gaps).

Table 8.1 presents a linear regression model that examines the association between the Facebook GGI and LinkedIn IT GGI, as well as the association between the LinkedIn overall GGI and LinkedIn IT GGI. As expected, the Facebook GGI has a positive and statistically significant association with LinkedIn gender gaps in IT ( $R^2=0.102$ ). Additionally, professional gender gaps on LinkedIn significantly predict the LinkedIn IT GGI, and explain a larger share of the variance therein than does the Facebook GGI ( $R^2=0.486$ ). In both models, however, adding in the Global Gender Gap (GGG) score – a female-to-male gender gap indicator that “benchmarks national gender gaps on economic, education, health and political criteria” (World Economic Forum, 2019, p. 8) and is comparable across the globe – further improves the model fit, although the coefficient on the GGG score is not statistically significant. Interestingly, however, the coefficient on the GGG score is negative, which suggests that once we control for general patterns of gender inequality in terms of digital access or professional gender inequality, less gender egalitarian countries show a tendency towards a greater share of women in IT. This could be interpreted in light of the ‘gender equality paradox’ described by Stoet & Geary (2018), through which quality of life pressures in less gender egalitarian countries might work to promote girls’ and women’s engagement in STEM fields that offer higher job security. Nevertheless, as this pattern is only observed in a handful of countries, it is difficult to estimate this effect with more statistical precision based on the data available.

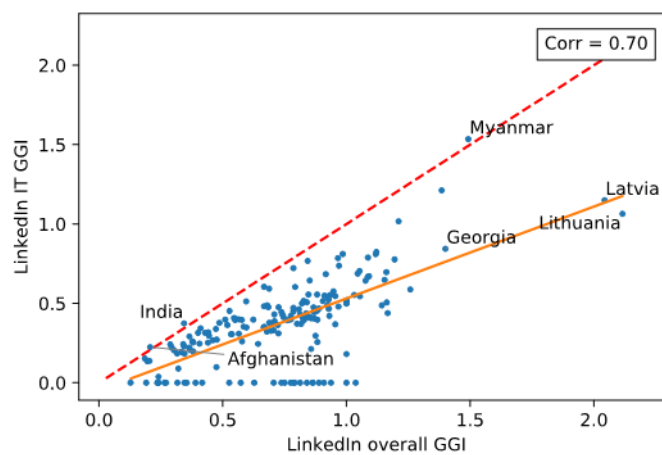
Figure 8.5. Scatter plot comparing low-level digital gender gaps as approximated by the FB GGI (horizontal axis) and high-skilled digital gender gaps as approximated by LinkedIn IT GGI (vertical axis)



Source: Authors own 2020, unpublished, using data from Facebook Marketing API<sup>9</sup> and LinkedIn Campaign Manager<sup>9</sup>

Note: Red dashed line shows  $y=x$  line, orange filled line shows fitted linear regression line.

Figure 8.6. Scatter plot comparing the overall LinkedIn GGI (horizontal axis) and the LinkedIn IT GGI (vertical axis)



Note: Red dashed line shows  $y=x$  line, orange filled line shows fitted linear regression line.

Source: Authors own 2020, unpublished, using data from LinkedIn Campaign Manager<sup>9</sup>

**Table 8.1. Summary of linear regression models predicting the LinkedIn IT GGI from gender inequality indicators**

Dependent variable: LinkedIn IT GGI				
	Facebook GGI		LinkedIn overall GGI	
	Model 1	Model 2	Model 1	Model 2
Intercept	0.147* (0.056)	0.197 (0.241)	-0.048 (0.035)	0.248 (0.151)
FB GGI	0.273*** (0.062)	0.382*** (0.076)		
LinkedIn overall GGI			0.579*** (0.044)	0.606*** (0.042)
GGG score		-0.138 (0.396)		-0.394 (0.233)
N	174	134	181	134
R <sup>2</sup>	0.102	0.229	0.486	0.646

\*p<0.05; \*\*p<0.01; \*\*\*p<0.001

\*p<0.05; \*\*p<0.01; \*\*\*p<0.001

Source: Authors own 2020, unpublished, using data from Facebook Marketing API<sup>9</sup>, LinkedIn Campaign Manager<sup>9</sup> and World Economic Forum (2019)

## Explaining gender gaps in IT industries

In this section, we explore which (gender-specific) development variables (e.g. GDP per capita, gender gaps in economic opportunity, as described in the data and analysis outline) can help explain variation in LinkedIn IT GGIs. Table 8.2 shows the correlation coefficients of the LinkedIn IT GGI with these variables. The LinkedIn IT GGI is most strongly correlated with gender gaps in professional/technical employment, followed by composite indicators of gender gaps in economic opportunity and educational attainment as well as

STEM tertiary education. Notably, the correlation of the LinkedIn IT GGI is stronger with gender-specific development variables than a variable of general economic development (GDP per capita). Additionally, gender gaps in tertiary education graduates in ICT are only weakly correlated with the LinkedIn IT GGI while gender gaps in STEM education are more strongly correlated with the LinkedIn IT GGI. Finally, gender gaps in educational attainment negatively correlate with tertiary education gender gaps in ICT and STEM, in line with the aforementioned 'gender equality paradox' (Stoet & Geary, 2018).

**Table 8.2. Correlations between LinkedIn IT GGI and various development and gender-specific development indicators**

	LinkedIn IT GGI	UNESCO ICT edu. GGI	UNESCO STEM edu GGI	ILO prof./tech. GGI	ILO labour part GGI	GDP per capita	GGG econ. opp. GGI	GGG edu. att. GGI
<b>LinkedIn IT GGI</b>	1.000	0.028	0.229	0.524	0.080	0.116	0.297	0.400
<b>UNESCO ICT edu. GGI</b>		1.000	0.668	-0.271	-0.419	-0.170	-0.305	-0.202
<b>UNESCO STEM edu. GGI</b>			1.000	-0.086	-0.415	-0.091	-0.259	-0.071
<b>ILO prof./tech. GGI</b>				1.000	0.242	0.150	0.553	0.602
<b>ILO labour part GGI</b>					1.000	0.125	0.810	0.120
<b>GDP per capita</b>						1.000	0.148	0.416
<b>GGG econ. opp. GGI</b>							1.000	0.174
<b>GGG edu. att. GGI</b>								1.000

Source: Authors own 2020, unpublished, using data from International Labour Organization (2019b; 2019c), LinkedIn Campaign Manager<sup>9</sup>, United Nations Educational, Scientific and Cultural Organization (2019), World Bank Group (2020) and World Economic Forum (2019)

Table 8.3 summarizes results from two linear regression models predicting the LinkedIn IT GGI. In the first model, we include the composite gender gaps linked to educational attainment and economic opportunity, which we observe explain

variation in gender inequality in IT on LinkedIn significantly, while GDP per capita does not. This suggests that gender-specific development factors may matter more than overall economic development in explaining variation in IT gender



gaps. On average, gender inequality in IT is lower in countries where women are more equal to men in terms of educational and economic opportunities. The first model indicates that 22 per cent of the variance in the LinkedIn IT GGI is explained by these three variables ( $R^2 = 0.216$ ).

The professional and technical workers' GGI, which is a component of the economic opportunity GGI, is most strongly correlated with the LinkedIn GGI

(Table 8.2). In Model 2 in Table 8.3 we replace the economic opportunity GGI from Model 1 with the professional and technical workers' GGI, and add the highly correlated field-specific STEM education GGI, and remove the weakly correlated GDP indicator. With these more specific indicators, we are able to increase the explained variation in the LinkedIn IT GGI by five percentage points from 21.6 to 26.8 per cent, but at the cost of more limited country coverage of 87 instead of 134.

**Table 8.3. Summary of linear regression models predicting the LinkedIn IT GGI from development indicators**

	Dependent variable: LinkedIn IT GGI	
	Model 1	Model 2
Intercept	-1.041*** (0.262)	-0.279 (0.620)
Economic opportunity GGI	0.489** (0.161)	
Educational attainment GGI	1.215*** (0.272)	0.401 (0.669)
GDP per capita (scaled by one million)	-0.683 (0.855)	
STEM education GGI		0.080 (0.103)
Professional and technical workers GGI		0.289*** (0.064)
N	134	87
R <sup>2</sup>	0.216	0.268

\*p<0.05; \*\*p<0.01; \*\*\*p<0.001

\*p<0.05; \*\*p<0.01; \*\*\*p<0.001

Source: Authors own 2020, unpublished, using data from International Labour Organization (2019b), LinkedIn Campaign Manager<sup>9</sup>, United Nations Educational, Scientific and Cultural Organization (2019), World Bank Group (2020) and World Economic Forum (2019)

## Validating LinkedIn gender gaps in IT

Figure 8.7 shows how the LinkedIn IT GGI compares to the ILO IT GGI. To make sure that the LinkedIn IT GGI matches the ILO IT GGI – which is based on ISIC divisions 58 to 63 in section J, 'Information and communication' – as well as possible, we now additionally include the following industries when computing the LinkedIn IT GGI: motion pictures and film, broadcast media, newspapers, media production, publishing, writing and editing, computer games and online media. We observe that for most countries, gender equality in IT is overestimated on LinkedIn relative to ILO. This is particularly the case for countries like Georgia, Latvia, Lithuania, Myanmar and Viet Nam, while it is the opposite for Guyana, for example. We further observe that the correlation between the two measures is in the expected direction but of modest magnitude (0.21). Selecting only those countries with high (above median) LinkedIn penetration – calculated as the number of LinkedIn users divided by the population aged 20+ in the country – does not significantly improve the correlation. This relatively weak correlation may be linked to the fact that industries and subcategories from LinkedIn are

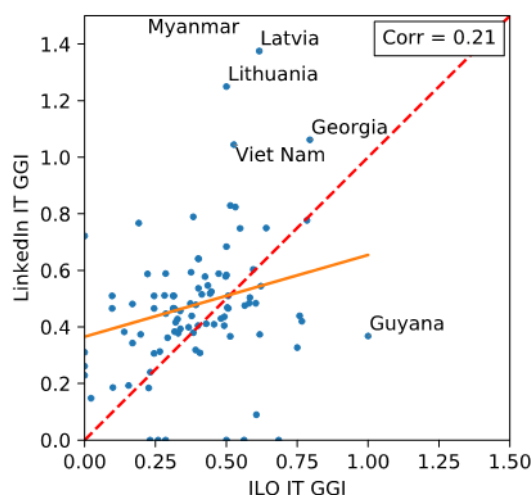
available on a more granular level than in more aggregated ILO statistics: LinkedIn includes specific industries, while the most granular ILO data is at the level of ISIC divisions. As such, the two measures may be mismatched in terms of the industries they capture. An advantage of the LinkedIn measure over that from ILO is that it has much better country coverage: while the LinkedIn measure is available for 181 countries, the ILO measure is available for only 107 countries.

## Discussion

In this paper, we have examined explanations for and variation in professional gender gaps in IT industries as captured in the population of LinkedIn users. We have firstly shown that gender gaps in IT are substantial worldwide. Although the IT sector is skewed towards men across all geographic and income regions on average, this skewness seems to be largest in regions such as Africa, and larger in low(er-middle) income countries. The IT gender gaps we find on LinkedIn are similar to those from Eurostat (2019), at least for European countries. Possible explanations that have been offered for these gender gaps are male



Figure 8.7. Scatter plot comparing the ILO IT GGI (horizontal axis) and the LinkedIn IT GGI (vertical axis)



Source: Authors own 2020, unpublished, using data from International Labour Organization (2019a), LinkedIn Campaign Manager<sup>9</sup> and World Bank Group & LinkedIn Corporation (2019)

dominance regarding usage frequency/patterns and time consumption of ICT, and male-directed work organization (Tømte, 2008; Valenduc & Vendramin, 2005).

While women are underrepresented in all IT industries considered, there is nevertheless some variation. There are almost four times more men than women in IT industries like computer hardware and networking, while there are almost twice as many men as women in the Internet and telecommunication industries. These results indicate that gender gaps are larger in computer-related IT industries than in IT industries related to communication and service. This may be because the latter require fewer digital-intensive skills than the former (Ramilo et al., 2005). Future research would benefit from further distinguishing between skills and positions within ICT to better understand where the gender gaps are largest and how these can be reduced most effectively.

Gender gaps in IT generally tend to be larger than those in other industries on LinkedIn and other online populations like Facebook. Gender-specific occupational and educational factors, and the former in particular, seem to matter more than economic development alone in explaining variation in IT gender gaps. This suggests that economic development may affect men and women differently in the professional area, and that developmental efforts that do not specifically address gender may not be sufficient

in themselves to reduce professional gender inequalities in the IT sector. We have also found that in countries where there is more gender equality in educational attainment and economic opportunity, there is more gender equality in IT on average. However, there are some exceptions to these results for a handful of countries (e.g. India, Myanmar). These findings suggest that in these countries either there are more women in IT or these women may be more likely to select themselves into having a LinkedIn profile. Future research would benefit from examining this paradox and its causes further.

Our validation exercise against ILO data showed that LinkedIn generally overestimates ILO on gender equality, particularly for Albania, Latvia, Lithuania, Myanmar and Viet Nam. A possible explanation for this overestimation is that the LinkedIn and ILO GGIs are mismatched because LinkedIn data report specific industries, while ILO data report industries (or economic activity according to ILO terminology) on the more aggregate ISIC division level.<sup>5</sup> This highlights the importance of and need for more granular data on occupational gender gaps in order to examine and validate gender gaps across different subdomains, such as industries or skills.

Our study has some limitations. First, LinkedIn does not provide representative data, particularly for countries where penetration is low (Zagheni & Weber, 2012), and as such may overestimate

gender equality; for example, when the LinkedIn population is disproportionately young, or when women are disproportionately more online than men. Additionally, LinkedIn is blocked in some countries (e.g. China, the Russian Federation, Islamic Republic of Iran), and although people in these countries can still sign up using a virtual private network, this may lead to biased audience estimates. We have focused on ratios rather than absolute numbers and contextualized potential biases in our analyses and interpretations to address these potential limitations. Second, an advantage of the LinkedIn data is that IT industries are measured on a granular level. However, this complicated our validation exercise and thus our understanding of how representative our measures of gender equality in IT industries are, particularly on a global scale. Nevertheless, positive and statistically significant correlations between gender gap indicators on economic opportunity and educational attainment point to the fact that the LinkedIn indicators we compute are measuring a phenomenon of interest, namely the representation of women in different IT industries.

## Conclusion

We have addressed gender inequality in IT using aggregate data on LinkedIn users, showing that

gender gaps in IT are substantial worldwide and vary across different IT industries. Although this analysis has focused on patterns in IT industries on a global scale, future research also drawing on LinkedIn's advertising platform could explore further stratification by skills, age and seniority. Additionally, we have shown that validation of the LinkedIn IT GGI from a non-representative sample is difficult because no ground truth data with the same granular level of IT industries is available globally. Because the LinkedIn data have great potential given the large country coverage and the ability to be queried frequently, further exercises to understand potential biases in the LinkedIn data would be welcome additions to the literature. Generally, IT sectors seem more gender equal in countries that are more gender egalitarian in terms of educational attainment and occupational opportunity. This also possibly amplifies the need for women in IT in less egalitarian countries to signal themselves on LinkedIn. This paradox should not be mistaken for gender equality in IT, and governments and policy-makers should aim to further reduce not only gender gaps in IT but more generally gender gaps in educational attainment and professional opportunity. Doing so would contribute to increasing the inclusion of women in the labour market and further reducing the gender wage gap.

## List of references

Borgonovi, F., Centurelli, R., Dernis, H., Grundke, R., Horvát, P., Jamet, S., Keese, M., Liebender, A.S., Marcolin, L., Rosenfeld, D., & Squicciarini, M. (2018). *Bridging the digital gender divide: Include, upskill, innovate*. Organisation for Economic Co-operation and Development. [www.oecd.org/internet/bridging-the-digital-gender-divide.pdf](http://www.oecd.org/internet/bridging-the-digital-gender-divide.pdf)

Davaki, K. (2018). *The underlying causes of the digital gender gap and possible solutions for enhanced digital inclusion of women and girls*. European Parliament's Policy Department for Citizens' Rights and Constitutional Affairs. [www.europarl.europa.eu/RegData/etudes/STUD/2018/604940/IPOL\\_STU\(2018\)604940\\_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2018/604940/IPOL_STU(2018)604940_EN.pdf)

Eurostat (2019). *Statistics explained — Science, technology and digital society: ICT specialists in employment* [Data set]. [https://ec.europa.eu/eurostat/statistics-explained/images/d/db/V4\\_ICT\\_specialists\\_in\\_employment\\_ICT2019.xlsx](https://ec.europa.eu/eurostat/statistics-explained/images/d/db/V4_ICT_specialists_in_employment_ICT2019.xlsx)

Fang, Z. (2016). *U.S. Patent No. 15/365,723*. Washington, D.C.: United States Patent and Trademark Office.

Fatehkia, M., Kashyap, R., & Weber, I. (2018). Using Facebook ad data to track the global digital gender gap. *World Development*, 107(July), 189–209. <https://doi.org/10.1016/j.worlddev.2018.03.007>

Haranko, K., Zagheni, E., Garimella, K., & Weber, I. (2018). Professional gender gaps across US cities. In Association for the Advancement of Artificial Intelligence, *Proceedings of the twelfth international AAAI conference on web and social media* (pp. 604–607). <https://aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17847>

International Labour Organization (2019a). *Employment by sex and economic activity – ISIC level 2 (thousands) – Annual* [Data set]. ILOSTAT database. [www.ilo.org/shinyapps/bulkexplorer44/?lang=en&segment=indicator&id=EMP\\_TEMP\\_SEX\\_EC2\\_NB\\_A](http://www.ilo.org/shinyapps/bulkexplorer44/?lang=en&segment=indicator&id=EMP_TEMP_SEX_EC2_NB_A)

International Labour Organization (2019b). *Employment by sex and occupation (thousands) – Annual* [Data set]. ILOSTAT database. [www.ilo.org/shinyapps/bulkexplorer55/?lang=en&segment=indicator&id=EMP\\_TEMP\\_SEX\\_OCU\\_NB\\_A](http://www.ilo.org/shinyapps/bulkexplorer55/?lang=en&segment=indicator&id=EMP_TEMP_SEX_OCU_NB_A)

International Labour Organization (2019c). *Labour force participation rate – ILO modelled estimates, July 2019* [Data set]. ILOSTAT database. [www.ilo.org/ilostat-files/Documents/Excel/MBI\\_15\\_EN.xlsx](http://www.ilo.org/ilostat-files/Documents/Excel/MBI_15_EN.xlsx)

International Telecommunication Union (2017). *ICT facts and figures 2017*. [www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf](http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf)

International Telecommunication Union (2019). *Measuring digital development: Facts and figures*. [www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2019.pdf](http://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2019.pdf)

Lipowiecka, J. & Kiriti-Nganga, T. (2016, September). *The Gender Dimensions of Services*. International Centre for Trade and Sustainable Development. <https://www.tralac.org/images/docs/10585/the-gender-dimensions-of-services-ictsd-september-2016.pdf>

Ramilo, C., Hafkin, N., & Jorge, S. (2005). *Gender equality and empowerment of women through ICT* (Women 2000 and Beyond, September). United Nations Division for the Advancement of Women. [www.unwomen.org/-/media/headquarters/media/publications/un/en/w200009dot05icte.pdf?la=en&vs=956](http://www.unwomen.org/-/media/headquarters/media/publications/un/en/w200009dot05icte.pdf?la=en&vs=956)

Roca, T. (2019, November). AI talent in the European labour market. LinkedIn Economic Graph. LinkedIn. <https://economicgraph.linkedin.com/content/dam/me/economicgraph/en-us/PDF/AI-Talent-in-the-European-Labour-Market.pdf>

State, B., Rodriguez, M., Helbing, D., & Zagheni, E. (2014). Migration of professionals to the U.S.: Evidence from LinkedIn data. In Aiello, L.M. & McFarland, D. (Eds.), *Proceedings of the Sixth International Conference on Social Informatics* (pp. 531–543). [https://doi.org/10.1007/978-3-319-13734-6\\_37](https://doi.org/10.1007/978-3-319-13734-6_37)

Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science*, 29(4), 581–593. <https://doi.org/10.1177/0956797617741719>

Tømte, C. (2008, June 2–3). ‘Return to gender’: Gender, ICT and education [Background paper]. OECD expert meeting hosted by the Norwegian Ministry of Education and Research, Oslo, Norway. <https://pdfs.semanticscholar.org/e1ff/3e4f510cbd8a7e12facc55b0a5e396e5c26f.pdf>

United Nations Department of Economic and Social Affairs (2008). *International Standard Industrial Classification of all economic activities (ISIC), Rev. 4*. [https://unstats.un.org/unsd/publication/seriesM/seriesm\\_4rev4e.pdf](https://unstats.un.org/unsd/publication/seriesM/seriesm_4rev4e.pdf)

United Nations Educational, Scientific and Cultural Organization (2019). *Education: Percentage of female graduates by field of study* [Data set]. <http://data.uis.unesco.org/index.aspx?queryid=164>

Valenduc, G., & Vendramin, P. (2005, June 9–10). *Work organisation and skills in ICT professions: the gender dimension* [Paper presentation]. Conference: ICT, the Knowledge Society and Changes in Work, Den Haag, the Netherlands. <http://ftu-namur.org/fichiers/DenHaag-SISWO-pvgv.pdf>

World Bank Group (2020). *GDP per capita, PPP (current international \$)*. [Data set]. World Bank Group. <http://api.worldbank.org/v2/en/indicator/NY.GDP.PCAP.PP.CD?downloadformat=csv>

World Bank Group & LinkedIn Corporation (2019). *LinkedIn industry to ISIC Rev. 4 industry mapping reference* [Data set]. World Bank Group. [https://development-data-hub-s3-public.s3.amazonaws.com/ddhfiles/144635/linkedin\\_to\\_isic\\_rev\\_4\\_industry\\_mapping\\_0.csv](https://development-data-hub-s3-public.s3.amazonaws.com/ddhfiles/144635/linkedin_to_isic_rev_4_industry_mapping_0.csv)

World Economic Forum (2019). *Global gender gap report 2020*. [http://www3.weforum.org/docs/WEF\\_GGGR\\_2020.pdf](http://www3.weforum.org/docs/WEF_GGGR_2020.pdf)

Zagheni, E., & Weber, I. (2012). You are where you e-mail: Using e-mail data to estimate international migration rates. In Association for Computing Machinery, *Proceedings of the 4th Annual ACM Web Science Conference* (pp. 348–351). <https://doi.org/10.1145/2380718.2380764>

Zhu, T. J., Fritzler, A. & Orlowski, J. (2018). *Data insights: Jobs, skills and migration trends methodology and validation results*. World Bank Group & LinkedIn. <http://documents.worldbank.org/curated/en/827991542143093021/pdf/132009-REVISED-v1-08218-WB-LinkedIn-Report-FINAL3-pxp.pdf>

## Endnotes

- <sup>1</sup> LinkedIn (2020). *Campaign manager*. <https://www.linkedin.com/campaignmanager/accounts>
- <sup>2</sup> LinkedIn (2020). *About LinkedIn*. <https://about.linkedin.com>
- <sup>3</sup> For the definition of this targeting criterium, see: LinkedIn (2020). *Targeting options for LinkedIn advertisements*. <https://www.linkedin.com/help/lms/answer/722/targeting-options-for-linkedin-advertisements?lang=en>
- <sup>4</sup> We refer to professional and technical employment as Major Groups 2 and 3 of the International Standard Classification of Occupations (ISCO-08), in correspondence with World Economic Forum (2019).
- <sup>5</sup> According to World Bank Group & LinkedIn Corporation (2019), ISIC division 62 ‘computer programming, consultancy and related activities’ includes the following LinkedIn industries: computer hardware, computer software, computer networking, semiconductors, computer and network security and wireless. ISIC division 63 ‘telecommunications’ includes the LinkedIn industries of Internet and telecommunications. The LinkedIn industry of information technology and services is contained in ISIC division 63 ‘information service activities’.
- <sup>6</sup> World Bank Group (2020). World Bank Country and Lending Groups. <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>
- <sup>7</sup> United Nations Department of Economic and Social Affairs Statistics Division (2020). Standard country or area codes for statistical use (M49). <https://unstats.un.org/unsd/methodology/m49/>
- <sup>8</sup> Further information on the Facebook GGI can be found at: Digital Gender Gaps Project (2020). *Using big data to measure global gender gaps in Internet and mobile access*. [www.digitalgendergaps.org](http://www.digitalgendergaps.org)