

# A large-scale analysis of Facebook's user-base and user engagement growth

YONAS MITIKE KASSA<sup>1,2,\*</sup>, RUBÉN CUEVAS<sup>2,3,\*</sup>, AND ÁNGEL CUEVAS<sup>2,\*</sup>

<sup>1</sup>IMDEA Networks Institute, Madrid, Spain

<sup>2</sup>Department of Telematics Engineering, Universidad Carlos III de Madrid, Leganés, Spain

<sup>3</sup>UC3M-Santander Big Data Institute, Madrid, Spain

\*e-mail: yonas.kassa@imdea.org, rcuevas@it.uc3m.es, acrumin@it.uc3m.es

Corresponding author: acrumin@it.uc3m.es

**ABSTRACT** Understanding the evolution of the user-base as well as user engagement of online services is critical not only for the service operators but also for customers, investors, and users. While we can find research works addressing this issue in online services such as Twitter, MySpace or Google+, such detailed analysis is missing for Facebook, which is currently the largest online social network.

This paper presents the first detailed study on the demographic and geographic composition and evolution of the user-base and user engagement in Facebook over a period of three years. To this end, we have implemented a measurement methodology that leverages the marketing API of Facebook to retrieve actual information about the number of total users and the number of daily active users across 230 countries and age groups ranging between 13 and 65+.

The conducted analysis reveals that Facebook is still growing and geographically expanding. Moreover, the growth pattern is heterogeneous across age groups, genders, and geographical regions. In particular, from a demography perspective, Facebook shows the lowest growth pattern among adolescents. Gender-based analysis showed that growth among men is still higher than the growth in women. Our geographical analysis reveals that while Facebook growth is slower in western countries, it presents fastest growth in developing countries mainly located in Africa and Central Asia, analyzing the penetration of these countries also shows that these countries are at earlier stages of Facebook penetration. Leveraging external socioeconomic datasets we also showed that this heterogeneous growth can be characterized by indicators such as availability and access to Internet, Facebook popularity, and factors related with population growth and gender inequality.

**INDEX TERMS** Social network services, Information services, Facebook, Technology social factors

## I. INTRODUCTION

IN the last few decades, research on adoption and growth of technologies has been a focal of interest across multiple disciplines including marketing, economics, and sociology. Innovation diffusion theory, the concerns-based adoption model, technology acceptance model, the united theory of acceptance and use of technology are among the popular theories that aim to categorize technology adopters and adoption decision processes [1], [2]. On the other hand, the birth of Internet has dramatically revolutionized innovations in different fields including lifestyle and communications creating new research disciplines such as computational social science and online social network analysis that leverage the digital traces of users.

Following the advent of Internet, in the last decade we

have witnessed a rapid proliferation of online services, especially Online Social Networks (OSNs) and social media platforms, spanning a large user-base in the order of hundreds of millions (Instagram, Twitter, Google+, Snapchat) or even billions (Facebook, YouTube, WeChat) of registered users. The main business drivers behind most of these services are marketing and advertising, based on the rich data they collect about their users. Therefore, having a large user-base is a key factor for the financial success of these online services [3]–[5]. Being aware of this, the now leading companies such as Facebook, Google, and Twitter, avoided generating revenues to solely focus on increasing the user-base in their initial days.

However, it is not enough to just have a large user-base. The above mentioned marketing and advertising mechanisms

are effective only if the engagement of the users is substantial. Therefore, online services must aim at having the largest possible number of users engaging with the service on a daily basis. Typical metrics used by investors and financial stakeholders to assess such level of engagement include the number of daily active users and the average daily active time users spend in the service.

The lucrative business associated with popular online services has generated a very competitive ecosystem, where new services appear constantly. This has led to a very dynamic ecosystem where users can easily migrate from one service to a similar one at any given time. Indeed, we have observed the demise of several online social networks, which were popular at some point in the last decade: Windows Live Spaces, Bebo, Friendster, Orkut, and Myspace [6].

In the described scenario, it is very important for operators, customers (e.g., advertisers using the online service as a marketing platform) and investors to monitor the health of the service through metrics such as the user-base or number of daily active users. Specifically, the temporal evolution of these metrics needs to be monitored to properly assess the health of an online service. This would help identify growth potentials or to anticipate the loss of users and/or the reduction of engagement so that investors and customers can make informed decisions with respect to the service. In addition, the online service operator needs to go one step further by identifying the reasons behind the observed trend in the evolution of user-base and user engagement. This would allow the service operator to reinforce the practices leading to positive trends and take corrective measures on negative trends before its users, investors, and customers lose their interest in the service.

The research community has understood the importance of this issue. We find several works analyzing the evolution of the user-base in MySpace, Twitter or Google+ [7]–[9] while just few works are able to analyze the evolution in user engagement [8], [10]. These works use, in general, crawling techniques to collect a sample of user profiles that help them estimate the overall user-base as well as its evolution. While these works have been enlightening and of high value, it is hard to assess the accuracy of their estimations due to the lack of ground truth data, which is proprietary data owned by the online service operator. Moreover, these techniques depend on the "friendliness" of the online service to be crawled. In particular, Facebook, the service we consider in this paper, presents enormous barriers for large-scale crawling techniques. Due to this, previous research attempts to investigate Facebook growth or discontinued use is limited to using a very small sample of users (usually confined to a small demographic or geographical area) [11]–[13]. Cannarella et al. [14] used a different approach by combining search queries and contagion epidemic models to study the growth evolution of Facebook's global user-base. This work concluded that the OSN will see a rapid decline and eventually die out. In fact, it wrongly predicted a reduction of up to 80% of Facebook's user-base between 2015 and 2017. Considering Facebook

as an epidemic, this model is based on the following three possible states of users: Susceptible, Infected, Recovered.

Other relevant theories include the network effect in technology adoption which states a network becomes more valuable to its users as number of users in a network increase [15], and diffusion of innovation theory which classifies adopters into innovators, early adopters, early majority, late majority, and laggards (very conservative and techno phobic groups) based on their degree of innovativeness and their willingness to adopt new technology involving five stages of persuasion offering distinct strategies to appeal to each category [2]. Rogers further argues that for an innovation to self-sustain it has to be widely adopted and reach a critical mass [2]. Even though these models have their own strengths and weaknesses in their applicability to OSNs [15], [16], discussing how Facebook fits these models is out of the scope of this paper.

In this paper we analyze, for first time, the demographic and geographic composition and evolution of Facebook's user-base and user engagement by taking datasets spanning three years and across different countries and demographic groups. We also characterize the observed growth across countries using external datasets from sources such as World Bank and CIA world factbook showing that observed heterogeneity can be explained by countries' existing socio-economic and demographic standing. To this end, we define a methodology which leverages the Facebook marketing API offered to advertisers to configure their ad campaigns. This API allows to define queries including a specific target population based on demographic (age and gender) and geographical parameters (e.g. country or region). The API returns (among other information) the number of total users (user-base) and daily active users (user engagement) that match the parameters defined in the query. This methodology overcomes the limitations of previous approaches since it is scalable and directly obtains the actual data provided by the service operator itself. Moreover, the ability to measure the user-base and user engagement slicing per country and age groups allows us to analyze potential factors affecting the evolution of these important variables. Even though leveraging this API has been shown to be an effective alternative source of statistical information about users of the OSN [17]–[20], re-purposing this data source should be taken with caution, such as doing repeated measurements, and taking median value, comparing with alternative sources, when available (see SI Appendix Supplementary Text 1 for details and recommendations).

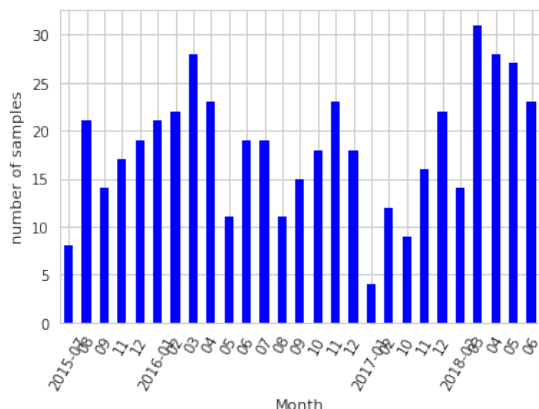
#### A. KEY INSIGHTS

Our analysis of Facebook's user-base and user engagement evolution over a period of three years and across 230 countries and age groups ranging between 13 and 65+ years has served to explain some fundamental aspects of Facebook growth in the last 3 years:

- Our Analysis indicates that Facebook is still growing (unlike previous studies that concluded otherwise) but at

July 2015			June 2018		
	DAU	TOT		DAU	TOT
Female	367644454	656408750	Female	608166893	1011000500
Male	377748364	799575020	Male	693520487	1280817500
Users in US	124634859	190430000	Users in US	162002253	242710000
Users in India	46390720	131810000	Users in India	150352982	290896000

**TABLE 1.** Counts of median Facebook total users (TOT) and daily active user (DAU) estimates, in July 2015 and June 2018 showing results for both genders, and results for US and India (countries having the largest user counts).



**FIGURE 1.** Distribution of days per month used in the analysis.

a very slow rate, where only half of its users are active on a daily basis.

- The evolution of user-base as well as user engagement is heterogeneous across age, gender and location. For instance, our analysis explains that the growth rate among adolescents is lower than other age groups (around 2.3 times smaller than the case of adults), while the growth rate of women is lower than for men (around 1.26 times smaller in user engagement growth and 1.1 times smaller in user-base growth). Moreover, Facebook shows a low to moderate growth in most analyzed countries. In particular, developed countries show a plateau in the growth trend, whereas the most important growth takes place in Africa and Central Asia.
- We have developed an explanatory model that considers socioeconomic factors related to Facebook growth across countries. This model shows that Facebook growth potential (measured by Facebook penetration) does not directly imply Facebook growth. In particular, (i) the user-base grows faster in areas having high urbanization rate, presenting a higher employment rate, and less infrastructure to access the Internet and thus Facebook (measured through Internet broadband access per 100 users), having higher gender inequality, and where the OSN is not among the top services in the country. (ii) the number of daily users presents a higher increase in countries showing a faster population growth, a decreasing unemployment rate, and where Facebook is not the most popular Internet service. The first three characteristics are representative to a high

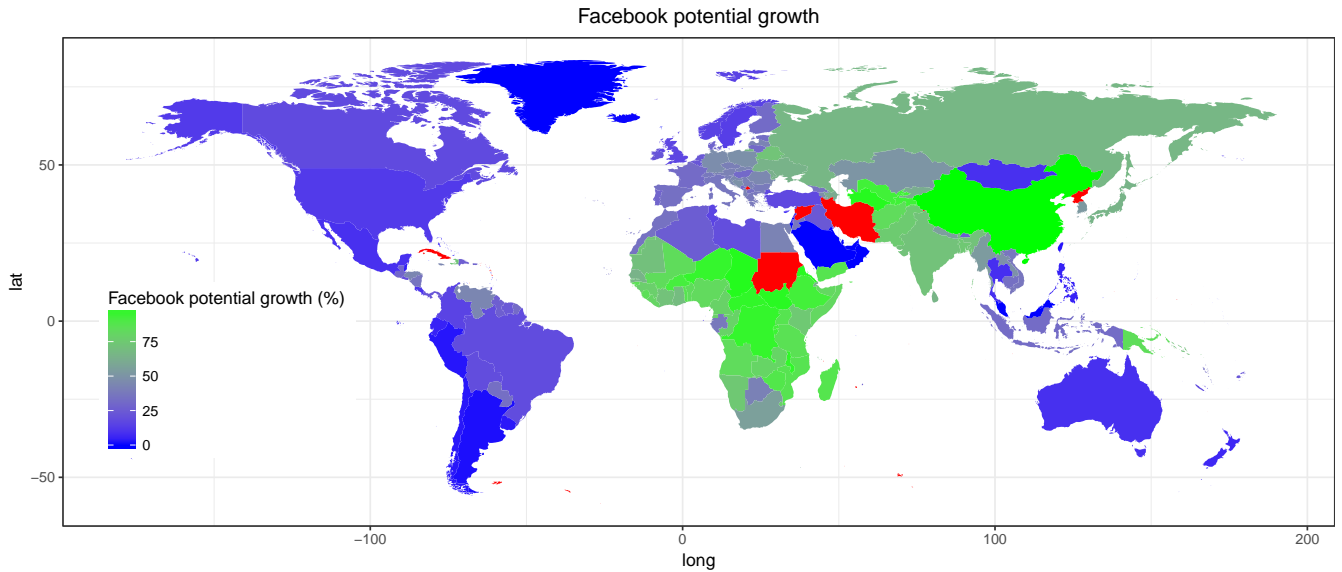
extent of emerging and pre-emerging countries with population resembling stage two of population pyramid [21], [22].

- The application of a random forest regression analysis on the three considered growth variability factors (i.e. geographic, age group, and gender) explains that geography is the variable influencing most of Facebook's growth in both user-base and engagement, followed by age group and gender.

## II. DATA COLLECTION TO MEASURE FACEBOOK USER-BASE AND USER ENGAGEMENT EVOLUTION

The Facebook dataset was collected directly from the OSN leveraging its marketing graph API. Intended for its customers, Facebook offers this feature-rich API [23] to enable them to reach a target audience defined by a range of demographic and behavioral targeting parameters. When queried with these pre-configured parameters, the API end point returns a JSON response which includes, among others, the number of total users, and the number of daily active users satisfying the targeted parameters. We query this API using geographic (i.e., countries or regions) and demographic (age and gender) parameters and without considering behavioral targeting parameters. Recently the research community has leveraged this API as an alternative way to extract actual datasets from the OSN [18], [24]–[31]. These works address different research questions than this paper.

Using this API we developed a distributed measurement system able to monitor the actual number of total users (user-base) and daily active users (user engagement) across ages and genders in all supported countries. The system is composed of a master program that handles authentication and load-balancing the data collection task among individual agent programs. Each agent contacts the Facebook API and starts querying the API, iterating over the list of targeting queries assigned to it by the master. Agents store the obtained responses in a central repository. Note that it is crucial to make sure that individual queries being parallelized do not contain overlapping target audiences. To preserve uniqueness of targeting queries, before assigning it to agents, the master recursively partitions our targeting parameters first based on countries, then based on gender (male vs female), finally each resulting query is partitioned into 53 groups based on age (from age = 13, the minimum age supported, through 65). We would like to note that age 65, which is the maximum possible targeting age, is actually an age group that includes



**FIGURE 2.** Global map of Facebook Potential growth: countries are colored based on their available room for Facebook growth ranging from dark blue (highest penetration hence less opportunity to grow) to light green (highest potential). Countries with no data are colored red.

users with age 65 and above, i.e. 65+. Moreover, Facebook offers three location types ("recent location", "home location", and "travel in") to target specific users. From these available options we have used the home location<sup>1</sup> of users to determine their geographical location since it is a more reliable way to identify "more permanent" location of users that use both mobile and desktop devices. As stated in their API documentation [23], Facebook uses a combination of techniques to reliably identify the "home location" of a user. These techniques include information based on IP address, "current city" in user's profile and from their friends' stated profile locations. In the the supplementary material we include a quantitative comparison analysis between home and recent location options that proves the former is a better choice.

Using the described measurement system we have obtained snapshots of the total number of users and daily active users for each age and gender group, in 230 available countries, extended over three years since July 2015 and collected in two periods between July 2015 to February 2017, and between October 2017 and June 2018. Every month in this dataset has a complete snapshot of at least 4 daily samples with a median of 19 days in a month. Figure 1 shows the number of days per month used in our analysis as we will see in Figure 7.

### III. ANALYSIS METHODOLOGY

Our goal is to measure the evolution of the user-base and user engagement in Facebook over the considered period of time. The number of total users obtained from our measurement is represented by the monthly active users variable reported by

<sup>1</sup>home location should not be confused with hometown.

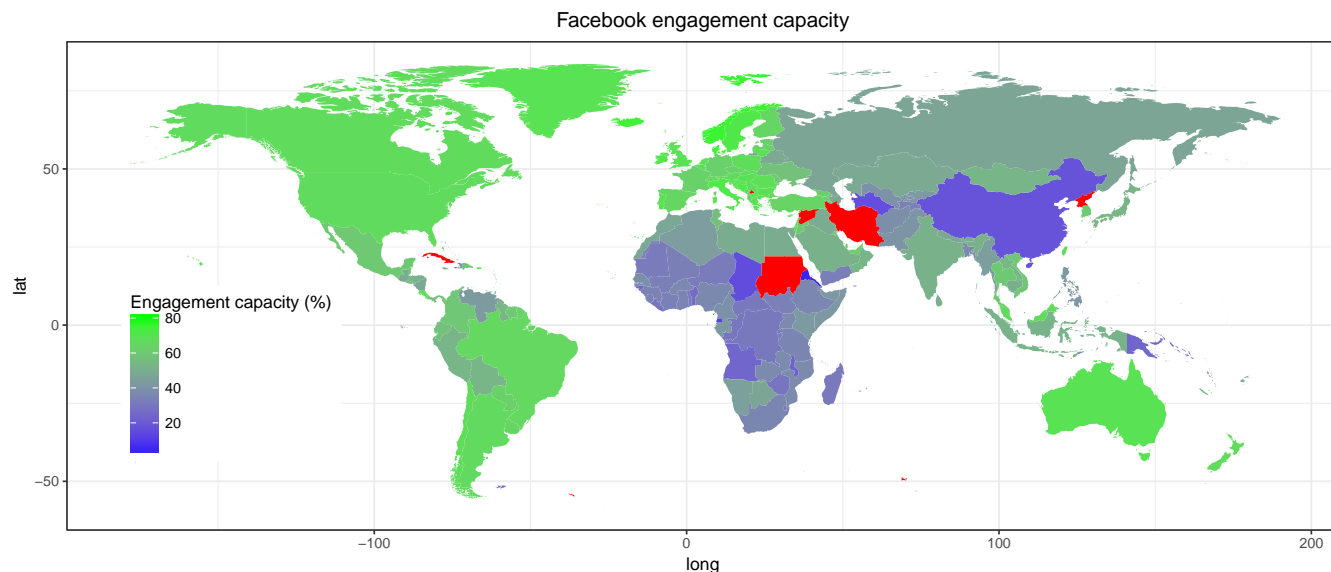
Facebook, we define this variable as the user-base. Moreover, the number of daily active users provides an aggregate measure of the extensively used user engagement. Therefore, we will consider these two variables for our analysis in the rest of the paper.

To compute the evolution of the proposed metrics over time, we use a reference value of the metric,  $x_0$ , which represents the median<sup>2</sup> number of Facebook total users (or daily active users) of a target population on the first month from our dataset. Then we compute the growth rate on a given month  $i$ , ( $U_i$ ), as the ratio of median number of added total users (daily active users) in that month and the reference value as:

$$U_i = \frac{x_i - x_0}{x_0} * 100$$

The above  $U_i$  values were calculated for both total users ( $U_{tot}$ ) and daily active users ( $U_{dau}$ ) of the target population. Note that the target population under study could be Facebook users of certain demographic or geographic group such as users per country, age group or gender. To derive the growth trends of the user-base and user engagement defined metrics, we analyze the temporal series of  $U_{tot}$  and  $U_{dau}$  as follows. First, a monthly temporal series of  $U_{tot}$  and  $U_{dau}$  is calculated based on the median values of each month. At the end of this step we obtain the evolution trend on monthly scale. Using this result we apply a regression model on the monthly growth values. In the model the growth value  $U_i$  is set as dependent variable on month number  $i$  which is calculated as the number of months since June 2015. At the end of this step we obtain model coefficients, which represent the

<sup>2</sup>In real world measurements median is more robust to outliers than mean. [32].



**FIGURE 3.** Global map of Facebook engagement capacity: countries are colored based on their observed engagement ranging from dark blue (least engagement among total users implying least engagement capacity) to light green (highest observed engagement). Countries with no data are colored red.

gradient of the regression line, which in turn indicates the rate of monthly growth of the metric as compared to the initial value  $x_0$ . For each metric,  $U_i$  indicates by how much the given value ( $U_{tot}$  or  $U_{dau}$ ) has changed as compared to  $x_0$ , for example a  $U_{tot} = 2.0$  tells that the number of total users has shown a 2% monthly increase as compared to  $x_0$ . An additional advantage of using a normalized metric  $U$  is that it allows a head-to-head comparison between targeted groups (e.g. comparison across demographic groups or countries).

#### IV. RESULTS

Using the above described methodologies and datasets we analyze the evolution of Facebook user-base and user engagement between July 2015 and June 2018. For reference Table 1 shows the median total users and DAU for first month and last months in our dataset. First we analyze a snapshot of our dataset in June 2018 to get a sense of Facebook's health status and potential growth capacity across countries and demographic groups. Then we analyze the evolution of the user-base and user engagement for the aggregated OSN as well as for different countries and age groups.

##### A. FACEBOOK'S GROWTH CAPACITY

According to our measurement, Facebook has 2.2 billion total users, and 1.3 billion daily active users as of June 2018, making it the most populous online community in the world. As a reference, the world population is 7.6 billion with China, India, and the US being the most populous countries with 1.39, 1.33, and 0.32 billion inhabitants, respectively.

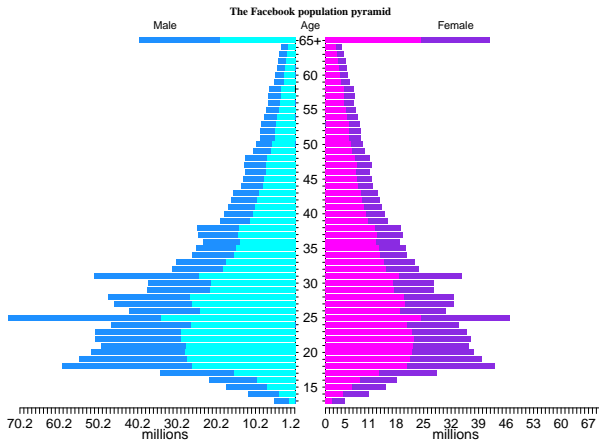
Figure 2 shows the growth capacity of Facebook in each country as of June 2018. Growth capacity measures the fraction of the population in a given country which is still

not on Facebook and is computed as  $1 - FB_p$ , where  $FB_p$  is Facebook penetration in each country.<sup>3</sup> Moreover, Figure 3, shows the ratio between number of daily active users and total users in each country as of June 2018. This metric captures engagement capacity of Facebook in different countries.

The results suggest that the OSN's growth potential lies in Asia and Africa. However, the currently existing engagement capacity in these geographical areas indicates that even though Facebook has a large room to grow in these regions, the engagement there may not reach the level of other areas with a larger penetration such as North America, South America or Europe. Note that Africa shows a higher engagement capacity than China due to the fact that Facebook is officially blocked in China, and relative the unavailability of other strongly competing services in Africa (e.g. WeChat of China has more than 900 million users).

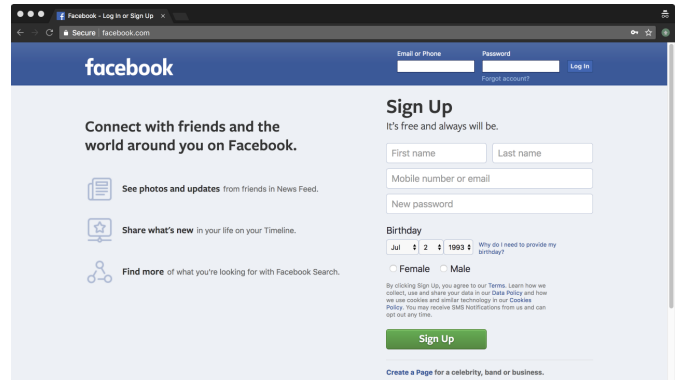
We now analyze the demographic population pyramid for the same dataset snapshot which is shown on Figure 4. Literature on demographic studies [22] classifies demographic pyramid of populations into three types: expansive pyramid, stationary pyramid, and constrictive pyramid. It is worth noting that Facebook's population pyramid does not exactly fit any of these three types. If we consider the number of births in Facebook as the population size at age 13 (minimum age to join the OSN) it can be characterized by a very low birth rate, and an increasing "immigration" (joining the site at a later age) up to the age of 19. If we take

<sup>3</sup>The Facebook penetration is computed as the ratio between the user-base of Facebook as of June 2018 and the country population as reported by US census bureau ([www.census.gov](http://www.census.gov)). Note that our methodology is subject to errors due to measurements errors by FB in computation of their user-base or DAU as well as errors in the computation of each country's population census over 13.



**FIGURE 4.** Global population pyramid of Facebook (median value of June 2018): The outer pyramid (darker color) represents demographic distribution of total users, inner pyramid (lighter color) shows daily active users.

the pyramid above the age of 20, it resembles an expansive population. In general, the smaller percentages of people in the younger age cohorts make the pyramid more similar to a Constrictive population (a pyramid constricted at the bottom with a lower percentage of younger people). A strict interpretation of this type of pyramid would suggest that the long-term survival of the social network may be questionable since the OSN is becoming less appealing to the younger generation, which will not use the service. An alternative explanation of the lower presence of young population at the bottom of the pyramid could be that the social network lacks features tailored to a younger population, and people join the site as they become older and find the features of the service more appealing. The stability of the pyramid shape, which shows a similar shape over the two analyzed years, suggests that the latter explanation is more plausible. If we consider the group  $G$  formed by the total users of age  $X$  in July 2015 and thus ages  $X + 1$  in July 2016,  $X + 2$  in July 2017, and  $X + 3$  in June 2018 for the first hypothesis to hold, the size of this group should (at most) remain the same. Instead, we observe a growth of such groups for ages ranging between 13 and 18, suggesting that as teenagers become older they find Facebook more interesting. Another interesting observation is that the pyramid does not follow a smooth trend at some age, typically showing a large spike on age 25, further investigation shows that this age group is related with the default year of birth (1993) put at Facebook registration page (Figure 5), which led us assume that many users just proceeded without modifying the default birth date set by the platform. Looking at the pyramid also indicates a male-biased disparity between genders in the platform as was recently revealed [18].



**FIGURE 5.** The default setting of Facebook registration page showing 1993 as a default year of birth.

## B. FACEBOOK GROWTH EVOLUTION

### 1) Overall Growth:

Let us start by considering the evolution of number of total users and daily active users over the analyzed period. Applying the previously described methodology, we conclude that Facebook is growing, contrary to the prediction of previous studies [14]. During our observation time, Facebook has grown from 1.45 billion users with 746 million daily active users in July 2015, to 2.3 billion users with 1.3 billion daily active users in June 2018. Table 1 shows comparative statistics between these two months.

These are doubtless impressive numbers. However, our analysis reveals that the growth rate in number of total users and active users is still male-biased indicating that gender inequality as seen through this social media is relentlessly increasing.

### 2) Growth Across Age Groups:

Based on Erikson’s stages of psycho-social development [33] we classify the age range into the following groups: adolescence (ages 13-19), early adulthood (ages 20-39), adulthood (ages 40-64), and maturity (ages 65+). Using this classification we analyze and compare the evolution of the Facebook user-base and user engagement in each stage. The computed metric values on each age group are presented on Figure 6. The overall observed trend is shown on figure 7.

The results show that Facebook growth rate is heterogeneous across age groups, with user-base growth of 0.98%, 1.78%, 2.30%, and 2.46% for adolescent, early adolescent, adulthood and maturity age groups, respectively. Having user engagement growth of 1.43%, 2.44%, 3.03%, and 3.27% for adolescent, early adolescent, adulthood and maturity age groups, respectively. This shows the growth among adolescents is significantly smaller, by a factor of (at least) two, than for adult groups. This result reinforces our conclusion from the demographic pyramid analysis, showing that Facebook seems to be less appealing for adolescents, that become more interested in the OSN as they become adults. Note that this might have implications from a marketing perspective since Facebook might not be the most appropriate venue to

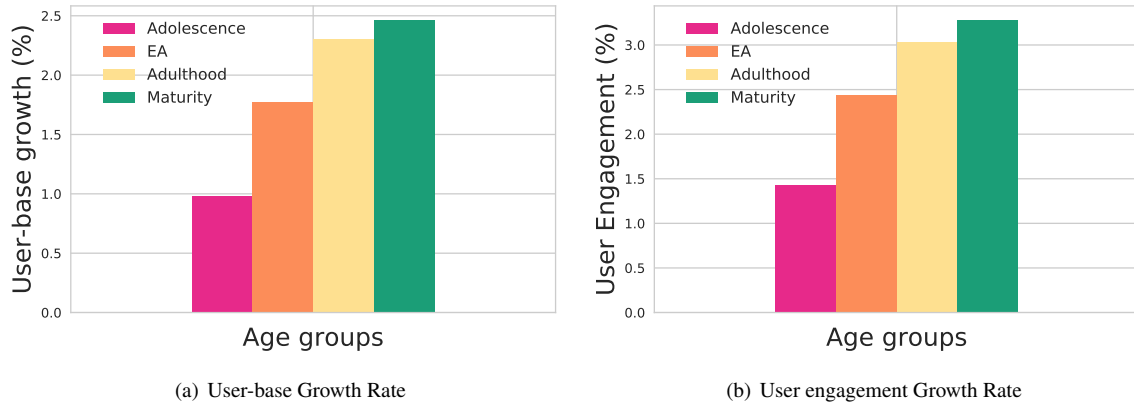


FIGURE 6. Growth rate of Facebook’s user-base (total users) and user engagement (daily active users) for different age groups: Adolescence (13-19), Early Adulthood (20-39), Adulthood (40-64) and Maturity (65+).

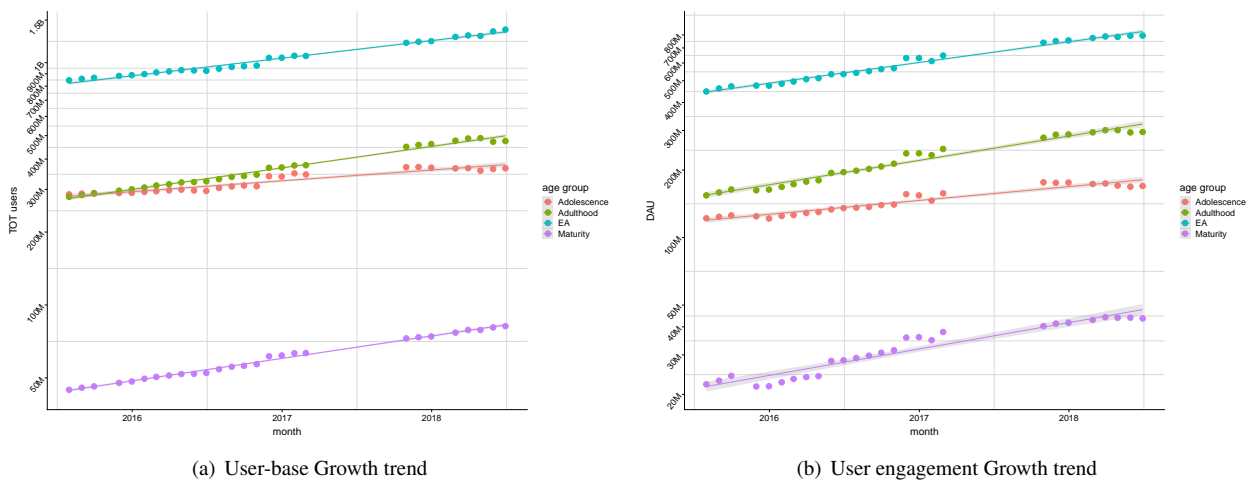


FIGURE 7. Monthly evolution trend of Facebook user-base and user engagement based on age groups. The left figure shows the evolution trend of user-base measured via total users, the right figure shows the trend in user engagement.

engage with adolescents. Given that Facebook is an OSN first popularized by US college students which could be considered as its innovators and early majority (according to Rogers technology adoption model [2]), Facebook has already gained momentum after astonishing growth over the past decade. We can observe that maturity age groups, the seemingly late majority and laggards, are joining the OSN at a high rate.

3) Growth across gender groups:

Here we discuss growth evolution based on gender. As a result for gender-wise comparison we study user-base and user engagement growth for men vs. women. The computed metric values for both genders are presented on Figure 8.

As shown in the figure, user engagement growth in both genders shows a higher growth rate than growth in user-base, suggesting that engagement in existing users grows as new users join the system. However, the growth in males is higher in both user-base (by a factor of 1.1) and user engagement (by a factor of 1.26) growth. Note that these

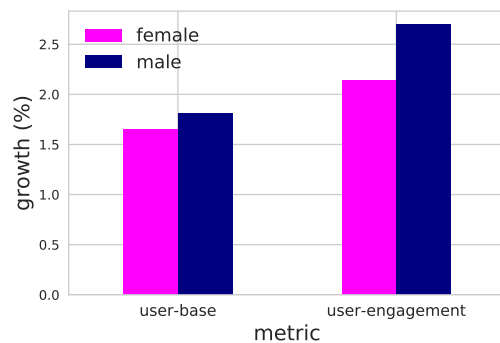


FIGURE 8. Growth rate of Facebook’s user-base (growth in total users) and user engagement (growth in daily active users) for male and female users on Facebook.

results demonstrate that the Facebook gender inequality is only increasing. Gender inequality in Facebook is thoroughly discussed in [18], the paper showed that gender inequality is associated with socioeconomic inequalities where higher

gender inequality is found in countries around Africa and southwest Asia.

#### 4) Growth across countries:

Here we use our described methodology to analyze the growth in the number of total users and daily active users across the 230 countries over the considered period of three years. Figure 9 shows the results.

A first look at the results reveals an overall higher growth in engagement than in total users, what can be interpreted as a sign of health suggesting Facebook's potential as a business platform (e.g., for marketing or advertising) is viable. If we consider individual countries, we observe that Facebook has slower growth and almost plateaued in most developed regions (US, Canada, EU, Scandinavian countries and Australia) whereas it is experiencing its most significant growth in Africa and Central Asia. This suggests that from social and business perspectives Facebook has reached an (almost) stable status in developed countries, where it has established itself as a de-facto social platform that connects a considerable fraction of the population. Instead, Facebook is currently spreading in some developing and underdeveloped geographical areas.

#### 5) Measuring Factors Influencing shift in usage/engagement

As we have seen above Facebook growth variability is demonstrated across geographic and demographic groups. Next, we collectively take age group, gender, and geographic dimensions and determine the influence of each dimension in growth variability. To measure influences we first calculate growth metric values for each group sliced per gender, country, and age group for both user engagement and user-base growth. We then apply random forest regression to identify the importance of each dimension in determining growth metric [32]. This methodology has been widely used in various disciplines [34], [35]. Our results show that in determining user-base growth, geographic dimension is the most influential variable following age group and gender with values 90.3%, 7.2%, and 2.5%, respectively. User engagement growth variability is also highly influenced by geographical growth followed by age group, and gender with values 71.6%, 22%, and 6.4%, respectively. Figure 5 in SI Appendix visually shows these values.

### C. EXTERNAL DATASETS TO CHARACTERIZE THE OBSERVED FACEBOOK EVOLUTION:

The previous visual interpretation is valuable to gather a first intuition. However, it does not reveal the underlying relations behind the observed heterogeneity of Facebook growth across countries. Our hypothesis is that this heterogeneity may be a reflection of the socioeconomic and demographic composition of countries. To explore the validity of this hypothesis, we leverage regression analysis techniques to measure the correlation between representative socioeconomic metrics and the growth of Facebook for different countries. To conduct our analysis we used the following datasets:

**(1) Demographic Metrics:** We leverage the population distribution dataset from census.gov<sup>4</sup>. This dataset includes the number of inhabitants of each sex and age country-wise. Using this dataset we aim to investigate the relation between Facebook's growth and its penetration in each country. We define Facebook penetration as the ratio between number of Facebook total users and the total population of a country. We also used this data source to get the population size and population density per sq. km. in each country. Moreover, datasets on urbanization rate and annual population growth were collected from The World Bank<sup>5</sup> to answer the following questions: "Does the sparsity of the population affect Facebook's growth?", "does rise in urbanization lead to a higher Facebook growth?, presuming a higher ICT infrastructure availability in urban areas". To measure relations with gender inequality we introduced ratios of male Facebook penetration and female Facebook penetration as gender-divide in a country.

**(2) Birth, Death, and Immigration Metrics:** We leverage datasets from CIA's world fact book<sup>6</sup> on birth and death rates per 1000 inhabitants of each country to investigate their relation with Facebook's growth. Moreover, we collect the net migration rate from the same dataset. One of the factors that contribute to population change in a country is migration. Even though the immediate impact of migration is insignificant to the overall population of a country, we can argue that migration is a net effect of seeking better social, economic and political opportunities not available in one's own country. Using this dataset we also aim to cover these unobserved factors.

**(3) Economic Metrics:** To account for economic metrics we leverage the GDP (gross domestic product) and GNI (gross national income) per capita which are primary indicators of a country's economic standing. We leverage the World Bank Report on GNI and GDP growth rate of world countries as of 2016<sup>7</sup>. The premise of using these datasets is to explore whether countries having a higher Facebook growth is correlated with economic prosperity of inhabitants.

**(4) Availability and Accessibility to Internet:** To explore the existence of correlations with the status of Internet in a country, we consider three signals collected from the World Bank: Internet affordability, Internet availability (Internet users per 100 inhabitants), and Quality of Internet (Fixed broadband Internet subscriptions per 100 inhabitants). Finally, we collected Alexa<sup>8</sup> ranking of Facebook in each country. Alexa Internet Rank is a relative measure of the popularity of an Internet service in a country as compared to other websites. Globally, Facebook ranks third right after Google and YouTube. Using this ranking, we aim to identify if Facebook is growing in countries where it is already popular or if it is expanding to new places where it is not.

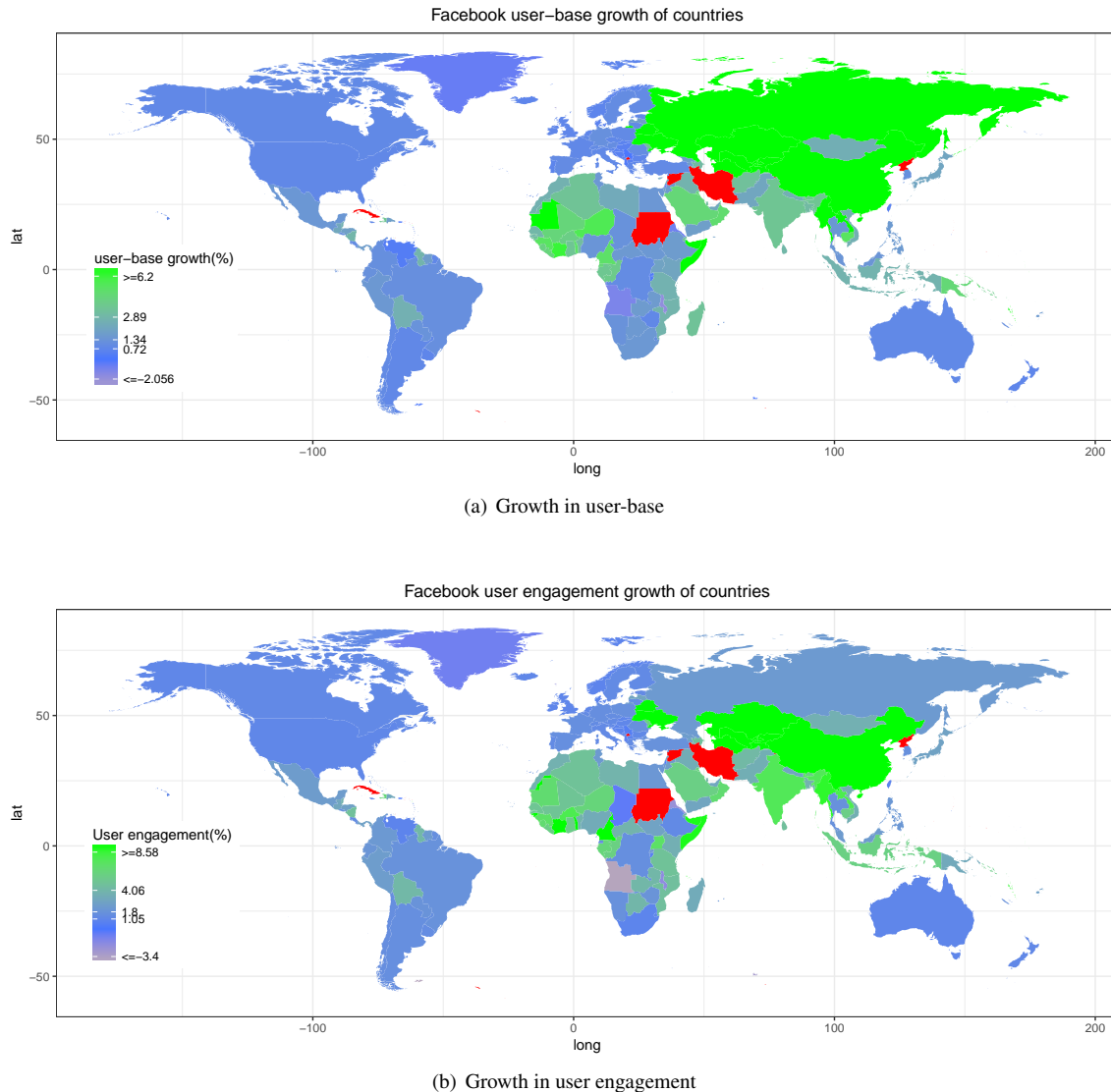
<sup>4</sup>www.census.gov

<sup>5</sup>data.worldbank.org

<sup>6</sup>cia.gov/library/publications/the-world-factbook/

<sup>7</sup>https://data.worldbank.org/indicator/

<sup>8</sup>www.alexa.com



**FIGURE 9.** Growth rate of number of total users (top) and number of daily active users (down). The range moves from lowest growth rate (yellow; not visible) to highest growth rate (light green).

### 1) Characterizing Facebook Evolution:

Here our goal is to model the multi-linear dependence relationship between the Facebook growth metrics of total users and daily active users per country computed using methodology in Section III and the explanatory variables described above. As discussed in Section III, growth metrics per country are calculated using the evolution dataset between July 2015 and June 2018. To this end, we use a linear regression model where multi-collinearity between variables is filtered using the variance inflation factor [32] (SI Appendix Table 1 reports the VIF values, and SI Appendix Figure 1 shows the correlation matrix between considered variables and growth metrics). The linear relationship models below show the dependence relation between evolution metrics and linear combination of most significant independent variables

for each metric. Table 2 shows the coefficients of resulting models.

$User\_base\_growth \sim broadband\_penetration + urbanization\_growth + Unemployment\_rate + FB\_alexa\_rank,$   
(with  $p$ -value  $< 3.91e-13$ , Adjusted  $R$ -squared = 0.63 and  $R$ -squared = 0.64)

$User\_engagement\_growth \sim birthrate + gender\_divide + Unemployment\_rate + FB\_alexa\_rank,$   
(with  $p$ -value  $< 2.83e-17$ , adjusted  $R$ -squared = 0.61 and  $R$ -squared = 0.62)

As it can be seen from the coefficients of determination, the relations between Facebook's growth metrics and the socioeconomic factors are relevant (adjusted  $R$ -square  $> 0.6$ ). Therefore, we can safely reject the null hypothesis that

Explanator	User_base	User_engagement
broadband_penetration	-0.33 ***	—
birthrate	—	0.38 ***
urbanization_growth	0.18 *	—
Unemployment_rate	-0.24 ***	-0.24 ***
gender_divide	—	0.24 ***
FB_alexa_rank	3.81 ***	3.35 ***

**TABLE 2.** Regression coefficients  
(Signif. codes: 0 '\*\*\*', 0.005 '\*\*', 0.05 '\*')

there is no relation between Facebook's growth metrics and socioeconomic factors.

First, the model suggests that Facebook has a higher user-base growth in areas with the following characteristics: where Facebook is not among the most popular Internet service, with higher urbanization growth, decreased unemployment rate, but with less Internet penetration (measured through broadband subscriptions per 100 people). Facebook seems to have noticed this potential; its recent "basic service" and the internet.org initiative efforts to expand Internet coverage to developing countries and areas with scarce connectivity seem to justify that [36].

On the other hand, user engagement growth shows a higher increase in countries characterized by: higher birthrate, decreased unemployment rate, but with higher gender inequality, and where the OSN is not in the top list of popular sites. The collective characteristics of countries showing the first three indicators represent a high extend of emerging or pre-emerging country with a stage two expansive pyramid [21], [22].

Even though correlation does not imply causation, and it is possible that other factors are responsible for the observed correlations, the examined attributes and associated results can be considered as an insight for further study into the observed phenomenon instead of as a prediction model of OSN growth in countries.

## V. CONCLUSION

This paper presents a novel methodology to monitor the growth rate of Facebook's user-base and user engagement with detailed demographic (per gender and ages ranging between 13 and 65+) and geographic granularity of 230 countries. The obtained results with this methodology are available at: [http://track.netcom.it.uc3m.es/fb\\_viz](http://track.netcom.it.uc3m.es/fb_viz).

This information is of high interest for Facebook's customers (mainly advertisers), social media analysts, and investors, which for first time have a system available to assess the composition and evolution of Facebook 'health' with such level of detail. Based on our experience, we have also included recommendations in using the API for research purposes (available in SI Appendix). At the same time the collected data available through the mentioned system has important value for researchers in multiple disciplines (computational social science, sociology, politics, etc). In particular, in this paper we present an initial analysis of the collected data over a period of three years, between July

2015 and June 2018, that report interesting results: Overall Facebook shows a growing trend. However, both Facebook user-base and engagement growth show gender bias with a higher growth rate in males than females (1.26 times smaller in user engagement growth and 1.1 times smaller in user-base growth). The growth pattern presents a clear heterogeneity across demographic groups and geographic regions. Facebook's growth rate is twice smaller among adolescents than adults. Facebook is reaching a plateau situation in developed regions of the world, while presenting the largest growth rate in developing countries mainly concentrated in Africa and Central Asia.

## ACKNOWLEDGMENT

Y.M.K. acknowledges funding from European H2020 project TYPES (grant number 653449). A.C. acknowledges funding from the Ministerio de Economía, Industria y Competitividad (Spain) and the European Social Fund (EU) through the Ramón Y Cajal Grant (RyC-2015-17732), and the Ministerio de Economía, Industria y Competitividad (Spain) through the project TEXEO (TEC2016-80339-R). R.C. acknowledges funding from the European H2020 project SMOOTH (grant number 786741), the 5GCity project (TEC2016-76795-C6-3-R) funded by the Spanish Ministry of Economy and Competitiveness and "La Caixa" Foundation, under agreement LCF/PR/MIT17/11820009.

## REFERENCES

- [1] Straub, E. T. Understanding technology adoption: Theory and future directions for informal learning. *Review of educational research* **79**, 625–649 (2009).
- [2] Rogers, E. M. *The diffusion of innovation* 5th edition (2003).
- [3] Dondio, P. Predicting stock market using online communities raw web traffic. In *Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 2012 IEEE/WIC/ACM International Conferences on, vol. 1, 230–237 (IEEE, 2012).
- [4] Parmy, O. Facebook closes 19 billion whatsapp deal. <http://newsroom.fb.com/news/2014/02/facebook-to-acquire-whatsapp/> (2014). [Online; accessed 08-Jan-2017].
- [5] Ben, T. Aol plans to sell or shut down bebo. <http://news.bbc.co.uk/2/hi/business/8606379.stm> (2010). [Online; accessed 08-Jan-2017].
- [6] Boyd, D. M. & Ellison, N. B. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication* **13**, 210–230 (2007). DOI 10.1111/j.1083-6101.2007.00393.x.
- [7] Rejaie, R., Torkjazi, M., Valafar, M. & Willinger, W. Sizing up online social networks. *IEEE network* **24** (2010).
- [8] Gonzalez, R., Cuevas, R., Motamedi, R., Rejaie, R. & Cuevas, A. Google+ or google-?: dissecting the evolution of the new osn in its first year. In *Proceedings of the 22nd international conference on World Wide Web*, 483–494 (ACM, 2013).
- [9] Schiöberg, D. et al. Tracing the birth of an osn: Social graph and profile analysis in google+. In *Proceedings of the 4th Annual ACM Web Science Conference, WebSci '12*, 265–274 (ACM, New York, NY, USA, 2012). DOI 10.1145/2380718.2380753.
- [10] Garcia, D., Mavrodiev, P. & Schweitzer, F. Social resilience in online communities: The autopsy of friendster. In *Proceedings of the First ACM Conference on Online Social Networks, COSN '13*, 39–50 (ACM, New York, NY, USA, 2013). DOI 10.1145/2512938.2512946.
- [11] Luqman, A., Cao, X., Ali, A., Masood, A. & Yu, L. Empirical investigation of facebook discontinues usage intentions based on sor paradigm. *Computers in Human Behavior* **70**, 544–555 (2017).
- [12] Barquero, B., Monreal, N., Ruiz-Munzón, N. & Serrano, L. Linking transmission with inquiry at university level through study and research

- paths: The case of forecasting facebook user growth. *International Journal of Research in Undergraduate Mathematics Education* **4**, 8–22 (2018).
- [13] Ravindran, T., Yeow Kuan, A. C. & Hoe Lian, D. G. Antecedents and effects of social network fatigue. *Journal of the Association for Information Science and Technology* **65**, 2306–2320 (2014).
- [14] Cannarella, J. & Spechler, J. A. Epidemiological modeling of online social network dynamics. (2014).
- [15] Hall, B. H. & Khan, B. Adoption of new technology. Tech. Rep., National bureau of economic research (2003).
- [16] Lyytinen, K. & Damsgaard, J. What's wrong with the diffusion of innovation theory? In *Working Conference on Diffusing Software Product and Process Innovations*, 173–190 (Springer, 2001).
- [17] Cabañas, J. G., Cuevas, Á. & Cuevas, R. Facebook use of sensitive data for advertising in europe. arXiv preprint arXiv:1802.05030 (2018).
- [18] Garcia, D. et al. Analyzing gender inequality through large-scale facebook advertising data. *Proceedings of the National Academy of Sciences* (2018). DOI 10.1073/pnas.1717781115.
- [19] Mejova, Y., Weber, I. & Fernandez-Luque, L. Online health monitoring using facebook advertisement audience estimates in the united states: Evaluation study. *JMIR public health and surveillance* **4** (2018).
- [20] Fatehkia, M., Kashyap, R. & Weber, I. Using facebook ad data to track the global digital gender gap. *World Development* **107**, 189–209 (2018).
- [21] Mahajan, V., Banga, K. & Gunther, R. The 86 percent solution: how to succeed in the biggest market opportunity of the next 50 years. (2005).
- [22] Korenjak-ĖGČerne, S., Kejžar, N. & Batagelj, V. Clustering of population pyramids. *Informatica* **32** (2008).
- [23] API, F. Facebook api for developers. <https://developers.facebook.com/docs/marketing-apis> (2016). [Online; accessed 08-Jan-2017].
- [24] Liu, Y., Kliman-Silver, C., Bell, R., Krishnamurthy, B. & Mislove, A. Measurement and analysis of osn ad auctions. In *Proceedings of the second ACM conference on Online social networks*, 139–150 (ACM, 2014).
- [25] Saez-Trumper, D., Liu, Y., Baeza-Yates, R., Krishnamurthy, B. & Mislove, A. Beyond cpm and cpc: Determining the value of users on osns. In *Proceedings of the second ACM conference on Online social networks*, 161–168 (ACM, 2014).
- [26] Korolova, A. Privacy violations using microtargeted ads: A case study. In *2010 IEEE International Conference on Data Mining Workshops*, 474–482 (IEEE, 2010).
- [27] Chen, T., Chaabane, A., Tournoux, P. U., Kaafar, M.-A. & Boreli, R. How much is too much? leveraging ads audience estimation to evaluate public profile uniqueness. In *International Symposium on Privacy Enhancing Technologies Symposium*, 225–244 (Springer, 2013).
- [28] Marciel, M., Gonzalez, J., Kassa, Y. M., Gonzalez, R. & Ahmed, M. The value of online users: Empirical evaluation of the price of personalized ads. In *Availability, Reliability and Security (ARES)*, 2016 11th International Conference on, 694–700 (IEEE, 2016).
- [29] González Cabañas, J., Cuevas, A., Cuevas, R. & Cuevas, R. Fdvt: Data valuation tool for facebook users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 3799–3809 (ACM, 2017).
- [30] Saha, K., Weber, I., Birnbaum, M. L. & De Choudhury, M. Characterizing awareness of schizophrenia among facebook users by leveraging facebook advertisement estimates. *Journal of medical Internet research* **19** (2017).
- [31] Araujo, M., Mejova, Y., Weber, I. & Benevenuto, F. Using facebook ads audiences for global lifestyle disease surveillance: Promises and limitations. In *Proceedings of the 2017 ACM on Web Science Conference, WebSci '17*, 253–257 (ACM, New York, NY, USA, 2017). DOI 10.1145/3091478.3091513.
- [32] Friedman, J., Hastie, T. & Tibshirani, R. *The elements of statistical learning* (Springer series in statistics Springer, Berlin, 2001).
- [33] Erikson, E. H. & Erikson, J. M. *The life cycle completed* (extended version) (WW Norton & Company, 1998).
- [34] Qi, Y. Random forest for bioinformatics. In *Ensemble machine learning*, 307–323 (Springer, 2012).
- [35] Casale, P., Pujol, O. & Radeva, P. Human activity recognition from accelerometer data using a wearable device. In *Iberian Conference on Pattern Recognition and Image Analysis*, 289–296 (Springer, 2011).
- [36] Sen, R. et al. On the free bridge across the digital divide: Assessing the quality of facebook's free basics service. In *Proceedings of the 2016 ACM on Internet Measurement Conference, IMC '16*, 127–133 (ACM, New York, NY, USA, 2016). DOI 10.1145/2987443.2987485.



Yonas also has a Master of Telematics Engineering degree from Universidad Carlos III de Madrid. During 2013 Yonas was an Erasmus exchange student at Universidad Carlos III of Madrid, where he completed his Masters thesis on the design of social-aware content distribution infrastructures for online social networks. Yonas has participated in the TMA shadow PC training, he has served as reviewer to journals and conferences such as IEEE Access, COMSNETS, AfriChi, WCNC, HOTPOST. His areas of interest include measurement and analysis of Internet traffic, online social networks, social media, privacy, and Information services in general.



DR. RUBÉN CUEVAS was born in Madrid, in 1981. He obtained his Ph.D. and M.Sc. in Telematics Engineering and M.Sc. in Telecommunications Engineering at University Carlos III of Madrid (Spain) in 2010, 2007 and 2005 respectively. Furthermore he received his M.Sc. in Network Planning and Management at Aalborg University (Denmark) in 2006. He is member of the Telematic Engineering Department and NETCOM research group since 2006, where he is currently Assistant Professor. Between January and December 2012 he was Courtesy Assistant Professor in the Computer and Information Science department at University of Oregon. In addition, from September 2008 until March 2009 he was intern in the Internet Scientific Group at Telefonica Research Lab Barcelona. He is coauthor of more than 70 papers in prestigious international journals and conferences such as IEEE Access, ACM CoNEXT, WWW, Usenix Security, ACM HotNets, IEEE Infocom, ACM CHI, IEEE/ACM TON, IEEE TPDS, CACM, PNAS, Nature Scientific Reports, PlosONE or Communications of the ACM. He has been the PI of 10 research projects funded by the EU H2020 and FP7 programs, the National Government of Spain and private companies and overall has participated in 24 research projects. Ruben's main research interests include Online Advertising, Web Transparency, Personalization and Privacy, Online Social Networks, and Internet Measurements. Ruben's research in file-sharing piracy, online social networks, online advertising fraud and web transparency has been featured in major international and national media such as The Financial Times, BBC, The Guardian, The Times, New Scientist, Wired, Corriere della Sera, O'Globo, Le Figaro, El Universal, El Pais, El Mundo, ABC, Cadena Ser, Cadena Cope, TVE, Antena3, La Sexta, etc.



DR. ÁNGEL CUEVAS received his M.Sc. in Telecommunication Engineering, M.Sc. in Telematics Engineering, and Ph.D. in Telematics Engineering from the Universidad Carlos III de Madrid in 2006, 2007, and 2011 respectively. He is currently a Ramón y Cajal Fellow (tenure-track Assistant Professor) in the Department of Telematic Engineering at Universidad Carlos III de Madrid and Adjunct Professor at Institut Mines-Telecom SudParis. His research interests focus on Internet Measurements, Web Transparency, Privacy and P2P Networks. He is co-author of more than 50 papers in prestigious international journals and conferences such as IEEE/ACM Transactions on Networking, ACM Transactions on Sensor Networks, Elsevier Computer Networks, IEEE Network, IEEE Communications Magazine, WWW, ACM CONEXT, ACM CHI, and ACM CoNEXT. He is recipient of the Best Paper Award in ACM MSWiM 2010.

...