Mitike Kassa, Yonas; Cuevas, Rubén; Cuevas, Ángel. (2018). A Large-Scale Analysis of Facebook's User-Base and User Engagement Growth. *IEEE Access*, v. 6, pp.: 78881-78891.

DOI: https://doi.org/10.1109/ACCESS.2018.2885458

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

Yonas Mitike Kassa
IMDEA Networks Institute
Universidad Carlos III de Madrid
Madrid, Spain
yonas.kassa@imdea.org

Ruben Cuevas
Universidad Carlos III de Madrid
Madrid, Spain
rcuevas@it.uc3m.es

Angel Cuevas
Universidad Carlos III de Madrid
Madrid, Spain
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 among 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.

Introduction

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 (Dondio 2012; Parmy 2014; Ben 2010). 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 (Boyd and Ellison 2007).

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 take 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+ (Rejaie et al. 2010; Gonzalez et al. 2013; Schiöberg et al. 2012) while just few works are able to analyze the evolution in user engagement (Gonzalez et al. 2013; Garcia, Mavrodiev, and Schweitzer 2013). 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 evolu-

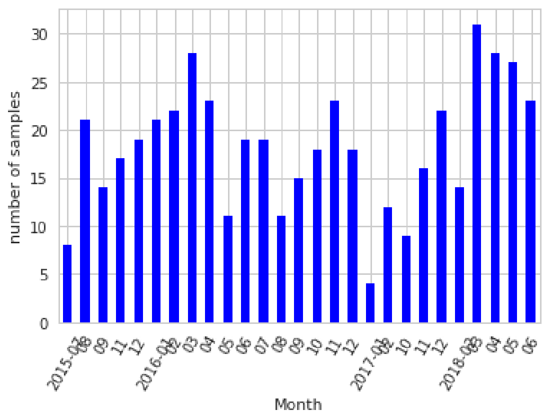


Figure 1: distribution of days per month used in the analysis.

tion. 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, we find only a previous paper studying the user-base evolution of Facebook using contagion epidemic models (Cannarella and Spechler 2014). 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.

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. 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 (country or region). The API returns (among other information) the total 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 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 effective and alternative source of statistical information about users of the OSN (Cabañas, Cuevas, and Cuevas 2018; Garcia et al. 2018; Mejova, Weber, and Fernandez-Luque 2018; Fatehkia, Kashyap, and Weber 2018), re-purposing this data sources should be taken with caution, such as doing repeated measurements, and taking median value, comparing with alternative sources, when available.

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 reveal the following findings:

- Facebook is still growing (unlike previous studies that concluded otherwise) but at 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, 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.
- Our analysis of socioeconomic factors related to Facebook growth across countries reveals 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 shows 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 (Mahajan, Banga, and Gunther 2005; Korenjak-Cerne, Kejžar, and Batagelj 2008);

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 (API 2016) to enable them 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 (Liu et al. 2014; Saez-Trumper et al. 2014; Korolova 2010; Chen et al. 2013; Marciel et al. 2016; González Cabañas et al. 2017; Saha et

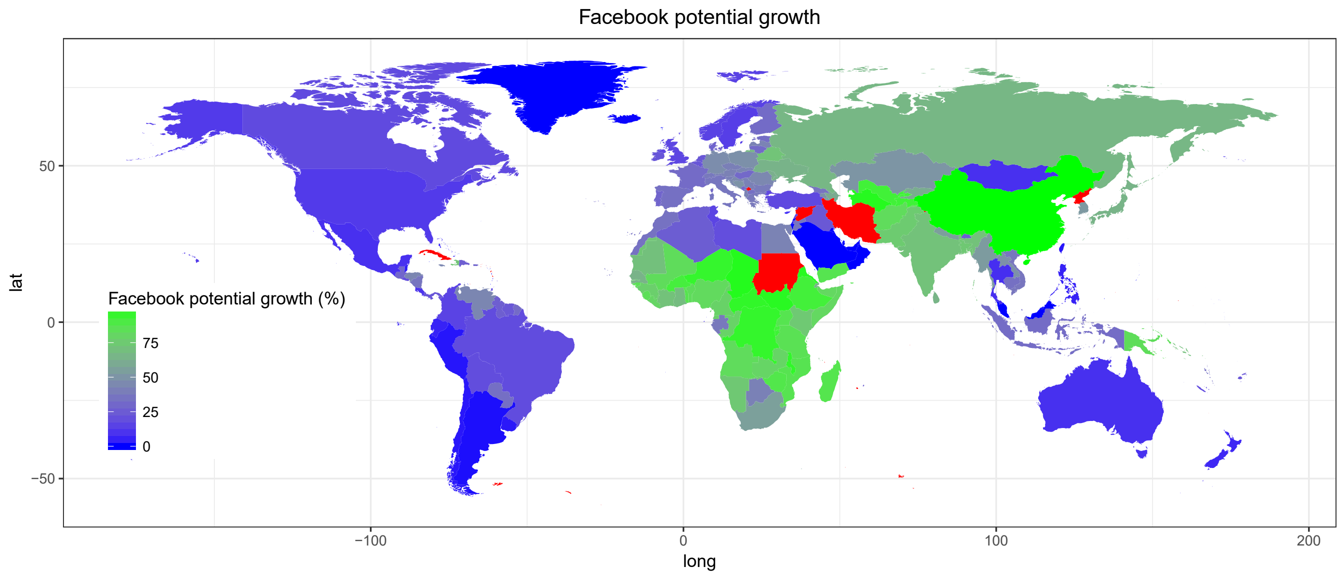


Figure 2: Global map of Facebook Potential growth: countries are colored based on their available room for Facebook growth ranging from dark blue (highest penetration hence less opportunity to grow) to light green (highest potential). Countries with no data are colored red.

al. 2017; Araujo et al. 2017; Garcia et al. 2018). 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 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 of users to determine their geographical location. As stated in their API documentation (API 2016), 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.

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 2016, and between October 2017 and June 2018. In this paper we consider a dataset including 27 months of data ranging between July 2015 and June 2018 with every month having 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.

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 by definition the user-base of Facebook. 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 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:

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

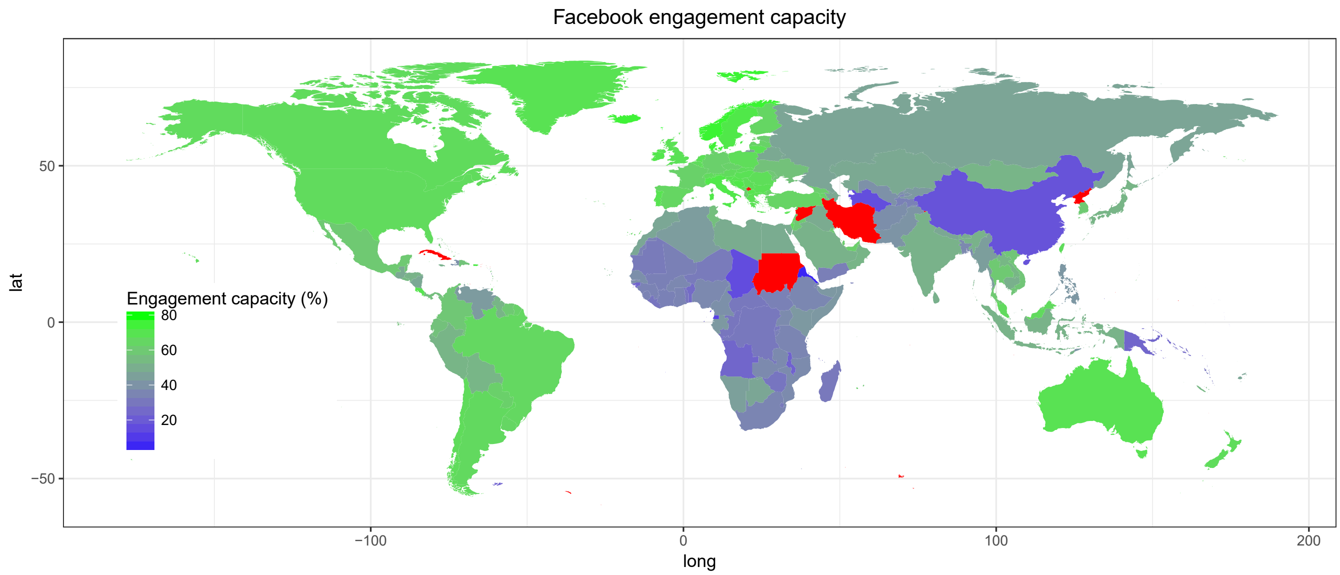


Figure 3: Global map of Facebook engagement capacity: countries are colored based on their observed engagement ranging from dark blue (least engagement among total users implying least engagement capacity) to light green (highest observed engagement). Countries with no data are colored red.

The above U_i values were calculated for both total users (U_{tot}) and daily active users (U_{dau}) of the target population.

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. At the end of this step we obtain model coefficients, which represent the 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).

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. First we analyze a snapshot of our dataset in June 2018 to get a sense of the 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.

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 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.[1] 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

[1]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).

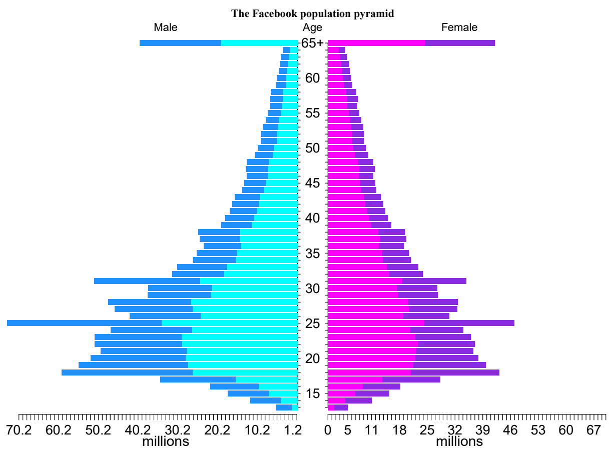


Figure 4: Global population pyramid of Facebook (median value of June 2018): The outer pyramid (darker color) represents demographic distribution of total users, inner pyramid (lighter color) shows daily active users.

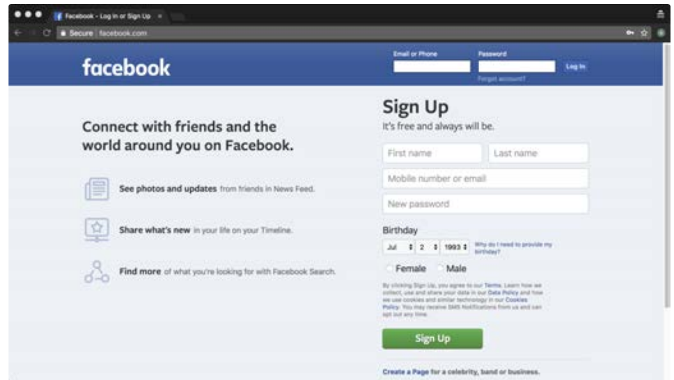


Figure 5: The default setting of Facebook registration page showing 1993 as a default year of birth.

the same dataset snapshot which is shown on Figure 4. Literature on demographic studies (Korenjak-Cerne, Kejžar, and Batagelj 2008) 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 the pyramid above 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 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 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 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 (Garcia et al. 2018).

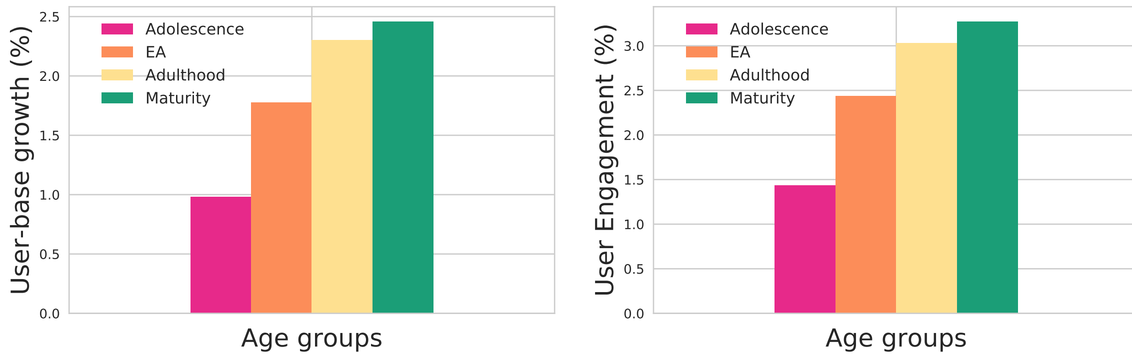
Facebook Growth Evolution

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 (Cannarella and Spechler 2014). 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.

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 digital inequality as seen through this social media is relentlessly increasing.

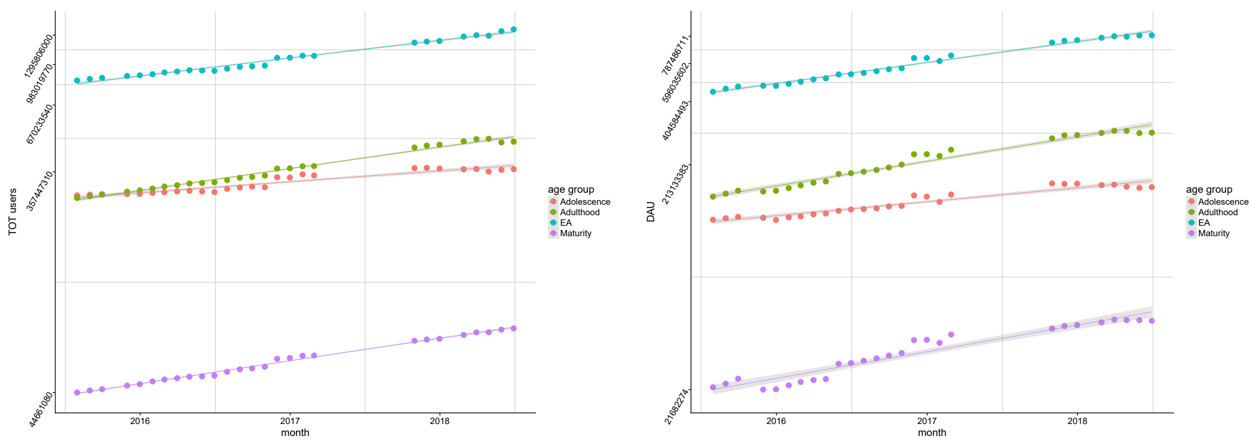
Growth Across Age Groups: Based on Erikson's stages of psycho-social development (Erikson and Erikson 1998) we classify the age range into the following groups: adolescence (ages 13-19), early adulthood (ages 20-39) , adulthood (ages 40-64) , and maturity (ages 65+). Using this classification we analyze 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 Face-



(a) User-base Growth Rate (b) User engagement Growth Rate

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



(a) User-base Growth trend (b) User engagement Growth trend

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.

book 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 most appropriate venue to engage with adolescents.

Growth across gender groups: In this section we discuss growth evolution based on gender. As a result for gender-wise comparison we study user-base and user engagement growth per group. The computed metric values for both genders are presented on Figure 6.

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 results demonstrate that the Facebook gender inequality, is only increasing. Gender inequality in Facebook is thoroughly discussed in (Garcia et al. 2018), the paper showed that gender inequality is associated with socioeconomic inequalities where higher gender inequality is found in countries around Africa and southwest Asia.

Growth across countries: In this subsection 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

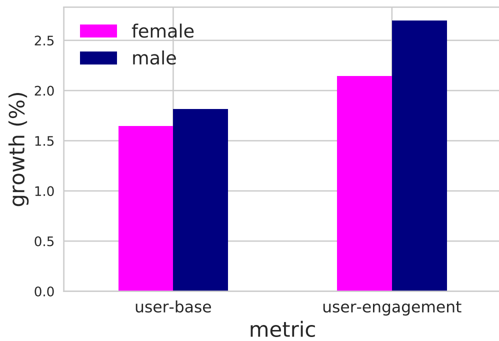


Figure 8: Growth rate of the Facebook's user-base (growth in total users) and user engagement (growth in daily active users) for male and female users on Facebook.

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.

Datasets to Characterize the Facebook Evolution: The previoius 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[2]. 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 the 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[3] to answer the following questions: "Does the sparsity of the population affects Facebook's growth?", "does rise in urbanization lead to a higher Facebook's 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[4] 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 [5]. The premise of using these datasets is to explore whether countries having a higher Facebook growth is correlated with economic prosperity of inhabitants.

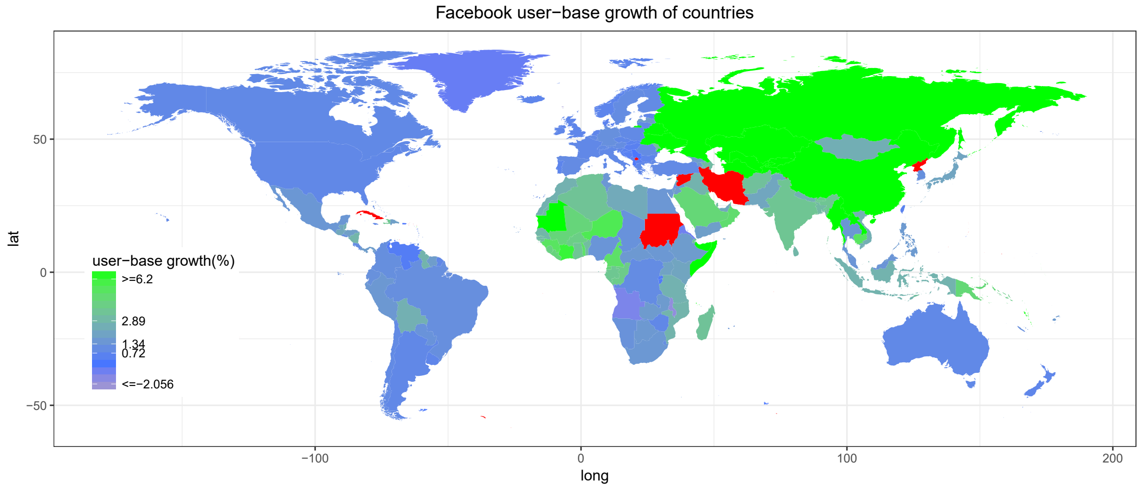
(4) Availability and Accessibility to Internet: To explore the existence of correlations with the status of Internet in a country, we consider three signals collected from the World Bank: Internet affordability, Internet availability (Internet users per 100 inhabitants), and Quality of Internet (Fixed broadband Internet subscriptions per 100 inhabitants). Finally we collected Alexa [6] 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 is it expanding to new places where it is not.

Characterizing Facebook Evolution: Here our goal is to model the multi-linear dependence relationship between the Facebook growth metrics (total users and daily active users) and the explanatory variables described above. To this end, we use a linear regression model where multi-colinearity between variables is filtered using the variance inflation factor (Friedman, Hastie, and Tibshirani 2001). The linear relationship models below show the dependence relation between evolution metrics and linear combination of most significant variables for each metric. Table 1 shows the coefficients of resulting models.

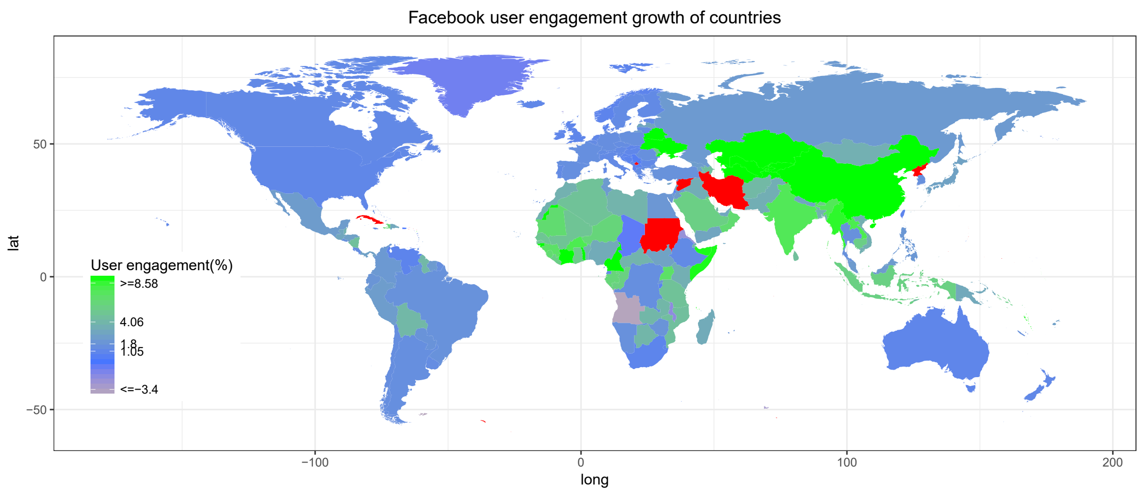
User_base_growth $\sim$ *broadband_penetration* + *urbanization_growth*+ *Unemployment_rate* + *FB_alexa_rank, (with p-value <3.91e-13, Adjusted R-squared = 0.63 and R-squared = 0.64)*

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)*

[2]www.census.gov
[3]data.worldbank.org
[4]cia.gov/library/publications/the-world-factbook/
[5]https://data.worldbank.org/indicator/
[6]www.alexa.com



(a) Growth in user-base



(b) Growth in user engagement

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).

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 1: Regression results
(Signif. codes: 0 '***', 0.005 '**', 0.05 '*')

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 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 (Sen et al. 2016).

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(Mahajan, Banga, and Gunther 2005; Korenjak-Cerne, Kejžar, and Batagelj 2008).

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.

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*.

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 evolution of Facebook 'health' with such level of detail. 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: (i) 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). (ii) 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.

Competing Interests:

The authors declare no competing interests.

References

[API 2016] API, F. 2016. Facebook api for developers. https://developers.facebook.com/docs/marketing-apis. [Online; accessed 08-Jan-2017].

[Araujo et al. 2017] Araujo, M.; Mejova, Y.; Weber, I.; and Benevenuto, F. 2017. 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. New York, NY, USA: ACM.

[Ben 2010] Ben, T. 2010. Aol plans to sell or shut down bebo. http://news.bbc.co.uk/2/hi/business/8606379.stm. [Online; accessed 08-Jan-2017].

[Boyd and Ellison 2007] Boyd, D. m., and Ellison, N. B. 2007. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication* 13(1):210–230.

[Cabañas, Cuevas, and Cuevas 2018] Cabañas, J. G.; Cuevas, Á.; and Cuevas, R. 2018. Facebook use of sensitive data for advertising in europe. *arXiv preprint arXiv:1802.05030*.

[Cannarella and Spechler 2014] Cannarella, J., and Spechler, J. A. 2014. Epidemiological modeling of online social network dynamics. *arXiv preprint arXiv:1401.4208*.

[Chen et al. 2013] Chen, T.; Chaabane, A.; Tournoux, P. U.; Kaafar, M.-A.; and Boreli, R. 2013. 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.

[Dondio 2012] Dondio, P. 2012. 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*, volume 1, 230–237. IEEE.

[Erikson and Erikson 1998] Erikson, E. H., and Erikson, J. M. 1998. *The life cycle completed (extended version)*. WW Norton & Company.

[Fatehkia, Kashyap, and Weber 2018] Fatehkia, M.; Kashyap, R.; and Weber, I. 2018. Using facebook ad data to track the global digital gender gap. *World Development* 107:189–209.

[Friedman, Hastie, and Tibshirani 2001] Friedman, J.; Hastie, T.; and Tibshirani, R. 2001. *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin.

[Garcia et al. 2018] Garcia, D.; Mitike Kassa, Y.; Cuevas, A.; Cebrian, M.; Moro, E.; Rahwan, I.; and Cuevas, R. 2018. Analyzing gender inequality through large-scale facebook advertising data. *Proceedings of the National Academy of Sciences*.

[Garcia, Mavrodiev, and Schweitzer 2013] Garcia, D.; Mavrodiev, P.; and Schweitzer, F. 2013. Social resilience in online communities: The autopsy of friendster. In *Proceedings of the First ACM Conference on Online Social Networks*, COSN '13, 39–50. New York, NY, USA: ACM.

[Gonzalez et al. 2013] Gonzalez, R.; Cuevas, R.; Motamedi, R.; Rejaie, R.; and Cuevas, A. 2013. 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.

[González Cabañas et al. 2017] González Cabañas, J.; Cuevas, A.; Cuevas, R.; and Cuevas, R. 2017. Fdvt: Data valuation tool for facebook users. In *Proceedings of the*

2017 CHI Conference on Human Factors in Computing Systems, 3799–3809. ACM.

[Korenjak-Cerne, Kejžar, and Batagelj 2008] Korenjak-Cerne, S.; Kejžar, N.; and Batagelj, V. 2008. Clustering of population pyramids. *Informatica* 32(2).

[Korolova 2010] Korolova, A. 2010. Privacy violations using microtargeted ads: A case study. In *2010 IEEE International Conference on Data Mining Workshops*, 474–482. IEEE.

[Liu et al. 2014] Liu, Y.; Kliman-Silver, C.; Bell, R.; Krishnamurthy, B.; and Mislove, A. 2014. Measurement and analysis of osn ad auctions. In *Proceedings of the second ACM conference on Online social networks*, 139–150. ACM.

[Mahajan, Banga, and Gunther 2005] Mahajan, V.; Banga, K.; and Gunther, R. 2005. The 86 percent solution: how to succeed in the biggest market opportunity of the next 50 years.

[Marciel et al. 2016] Marciel, M.; Gonzalez, J.; Kassa, Y. M.; Gonzalez, R.; and Ahmed, M. 2016. 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.

[Mejova, Weber, and Fernandez-Luque 2018] Mejova, Y.; Weber, I.; and Fernandez-Luque, L. 2018. Online health monitoring using facebook advertisement audience estimates in the united states: Evaluation study. *JMIR public health and surveillance* 4(1).

[Parmy 2014] Parmy, O. 2014. Facebook closes 19 billion whatsapp deal. http://newsroom.fb.com/news/2014/02/facebook-to-acquire-whatsapp/. [Online; accessed 08-Jan-2017].

[Rejaie et al. 2010] Rejaie, R.; Torkjazi, M.; Valafar, M.; and Willinger, W. 2010. Sizing up online social networks. *IEEE network* 24(5).

[Saez-Trumper et al. 2014] Saez-Trumper, D.; Liu, Y.; Baeza-Yates, R.; Krishnamurthy, B.; and Mislove, A. 2014. 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.

[Saha et al. 2017] Saha, K.; Weber, I.; Birnbaum, M. L.; and De Choudhury, M. 2017. Characterizing awareness of schizophrenia among facebook users by leveraging facebook advertisement estimates. *Journal of medical Internet research* 19(5).

[Schiöberg et al. 2012] Schiöberg, D.; Schneider, F.; Schiöberg, H.; Schmid, S.; Uhlig, S.; and Feldmann, A. 2012. 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. New York, NY, USA: ACM.

[Sen et al. 2016] Sen, R.; Pirzada, H. A.; Phokeer, A.; Farooq, Z. A.; Sengupta, S.; Choffnes, D.; and Gummadi, K. P. 2016. 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. New York, NY, USA: ACM.