Analyzing gender inequality through large-scale Facebook advertising data - Supplementary Information

David Garcia,[1,2] Yonas Mitike Kassa,[3,4] Angel Cuevas,[4]
Manuel Cebrian,[5,6] Esteban Moro,[5,7] Iyad Rahwan,[5,8*] Ruben Cuevas[4*]

[1]Complexity Science Hub Vienna, 1080 Vienna, Austria;
[2]Section for Science of Complex Systems,
Center for Medical Statistics, Informatics, and Intelligent Systems,
Medical University of Vienna, 1090 Vienna, Austria;
[3]IMDEA Networks Institute, 28918 Leganés, Spain;
[4]Department of Telematic Engineering, Universidad Carlos III de Madrid,
28911 Leganés, Spain;
[5]The Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139;
[6]Data61, Commonwealth Scientific and Industrial Research Organisation,
3008 Melbourne, Australia;
[7]Grupo Interdisciplinar de Sistemas Complejos, Department of Mathematics,
Universidad Carlos III de Madrid, 28911 Leganés, Spain;
[8]Institute for Data, Systems & Society,
Massachusetts Institute of Technology, Cambridge, MA 02139
[*]To whom correspondence should be addressed;
E-mail: irahwan@mit.edu, rcuevas@it.uc3m.es

June 1, 2018

www.pnas.org/cgi/doi/10.1073/pnas.1717781115

Supplementary Table 1

Gender	Facebook accounts	Median DAU	Total population
Female	646,953,680	375,148,380	2,422,856,560
Male	800,026,950	391,673,365	2,485,459,305
Both	1,446,980,630	766,821,745	4,908,315,865

Table 1: Total counts of Facebook accounts, DAU estimates, and total population covered by the dataset.

Supplementary Text 1 - Validating the FGD

We validate the measurement of the FGD against three reference datasets:

1. **Global Web Index (GWI).** We use the survey responses of the Global Web Index [1] panel during the period of our study (the two last quarters of 2015 and the two first quarters of 2016). For this period, the GWI contains responses from 99,338 panelists in 34 countries, providing rescaled estimates of survey responses that generalize to the population as a whole. We take the response to the question about the frequency of use of Facebook, calculating the weighted fraction of respondents for each gender and age segment that report to use Facebook daily or more than once a day. Furthermore, we repeat the analysis for other social media (Whatsapp, LinkedIn, Twitter, Instagram, and YouTube), calculating gender divide values outside Facebook.

2. **Pew Research Center Spring 2016 Global Attitudes (Pew Global).** This dataset includes responses from 23,462 panelists in 19 countries abd can be found online [2]. We use the positive answer rate to question 82 "Do you ever use online social networking sites like Facebook, Twitter...?" as way to measure the penetration of SNS in general. We take the respondent weights reported in the dataset to rescale the frequency of positive responses taking into account the self-reported gender of survey respondents.

3. **Pew Research Center Internet & Technology, March 7-April 4, 2016 (Pew US).** This US questionnaire [3] includes questions about Facebook use in particular (act135, "Do you ever use the internet or a mobile app to use Facebook?"), as well as gender and age data from 1,601 respondents. We use the respondent weights to compute Facebook use rates across gender and age groups.

[1] `https://www.globalwebindex.net/`
[2] `http://www.pewglobal.org/dataset/spring-2016-survey-data/`
[3] `http://www.pewinternet.org/dataset/march-2016-libraries/`

Comparison of Facebook penetration estimates

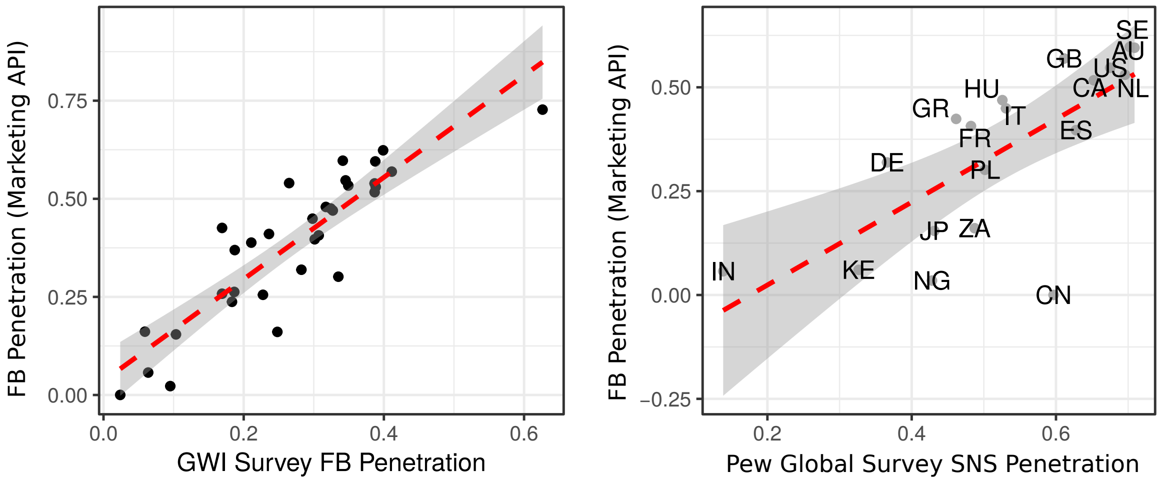

Figure 1: Total Facebook penetration for both genders as reported by the marketing API versus Facebook penetration as estimated in the GWI survey (left) and penetration of all Social Networking Sites in Pew Global survey (right). The red dashed line shows a linear regression profile, with its prediction standard errors in the shaded area.

The left panel of Fig. 1 shows the relationship between the total Facebook penetration for both genders as measured by us through the marketing API versus the value estimated from the GWI survey. There is a high positive correlation between both measurements of Facebook Penetration (Pearson 0.89 , CI $[0.78, 0.94]$, Spearman 0.86, CI $[0.72, 0.94]$). The right panel of Fig. 1 shows the same evaluation against the Pew Global survey. The correlation between both measurements is positive and high (Pearson: 0.71, CI $[0.39, 0.88]$, Spearman: 0.77, CI $[0.43, 0.94]$).

The left panel of Fig. 2 shows the comparison between estimates based on GWI and Pew Global survey data. The correlation is also positive and significant, even though samples are of limited size (Pearson: 0.63, CI $[0.17, 0.86]$, Spearman: 0.7, CI $[0.38, 0.91]$) Both in the right panel of Fig. 1 and in left panel of Fig. 2, it can be seen that China is a clear outlier. This stems from the difference between comparing Facebook penetration versus penetration for

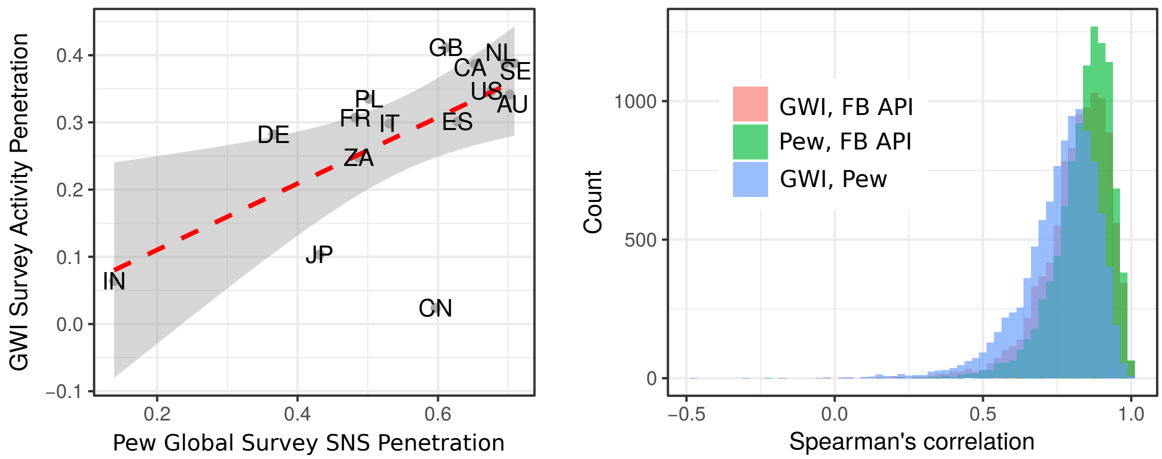

Figure 2: Left: Comparison of GWI Facebook penetration estimate and Pew penetration of all SNS. The red dashed line shows a linear regression profile, with its prediction standard errors in the shaded area. Right: Bootstrapping distributions of Spearman's correlation coefficient between all pairs of penetration measurements.

Social Networking Sites in general, which is the precise question of the Pew survey. If we focus on the rest of countries, where we can expect a priori that Facebook is more representative of social media in general, the correlation reaches higher values when comparing to the Facebook marketing API (Pearson: 0.84, CI $[0.62, 0.94]$, Spearman: 0.87, CI $[0.66, 0.96]$) and to the GWI survey data (Pearson 0.84, CI $[0.55, 0.95]$, Spearman 0.78, CI $[0.43, 0.94]$). To make a fair comparison, we should not include China when comparing Facebook penetration with penetration in SNS in general.

When we focus only on the set of countries available in all three datasets, we can compare the correlation of each pair of data sources to assess the validity of the Facebook API data in terms of Facebook penetration. In this smaller sample, the Facebook penetration calculated through the marketing API is highly correlated with both the Pew Global survey values (Pearson: 0.87, CI $[0.64, 0.96]$, Spearman: 0.86, CI $[0.59, 0.97]$) and with the GWI survey (Pearson: 0.89, CI $[0.68, 0.96]$, Spearman: 0.83, CI $[0.53, 0.97]$). The point estimates of these two values are higher than the correlation between the Pew Global and GWI datasets, as reported above.

Nevertheless, these differences are not significant, as confidence intervals overlap and bootstrap sampling shows that estimates are indistinguishable (Fig. 1). From this analysis we conclude that the estimate of Facebook penetration from the marketing API has comparable quality to the values reported in the high-quality, representative surveys of GWI and the Pew Research Center.

Facebook gender divide estimates

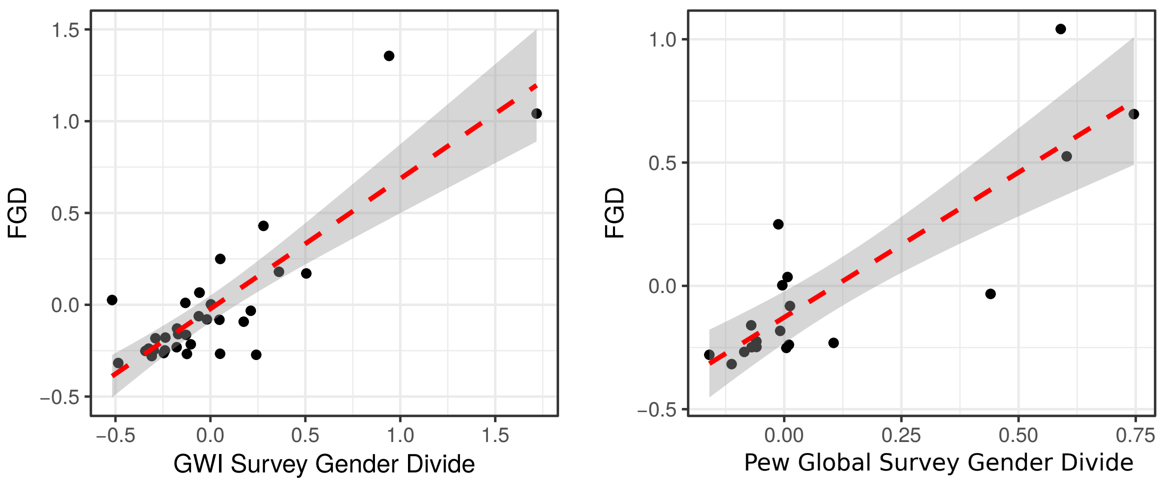

Figure 3: Measurement of the FGD in the API versus estimates using GWI data (left) and a gender divide for all SNS in PEW (right). The red dashed line shows a linear regression profile, with its prediction standard errors in the shaded area.

We calculated surrogates of the FGD based on the GWI and Pew global survey datasets. The left panel of Fig. 3 shows the comparison of the Facebook marketing API estimate with the same estimate using GWI data, revealing high positive correlation (Pearson: 0.83, CI $[0.68, 0.91]$, Spearman: 0.63, CI $[0.27, 0.87]$). The right panel of Fig. 3 shows the comparison of the Facebook marketing API estimate with Pew survey data for all SNS (including China), also revealing high positive correlation (Pearson: 0.85, CI $[0.65, 0.94]$, Spearman: 0.74, CI $[0.35, 0.91]$).

A comparison between estimates of the FGD using GWI versus using Pew data is shown on the left panel of Fig. 4, also revealing positive correlations (Pearson: 0.85, CI $[0.6, 0.95]$,

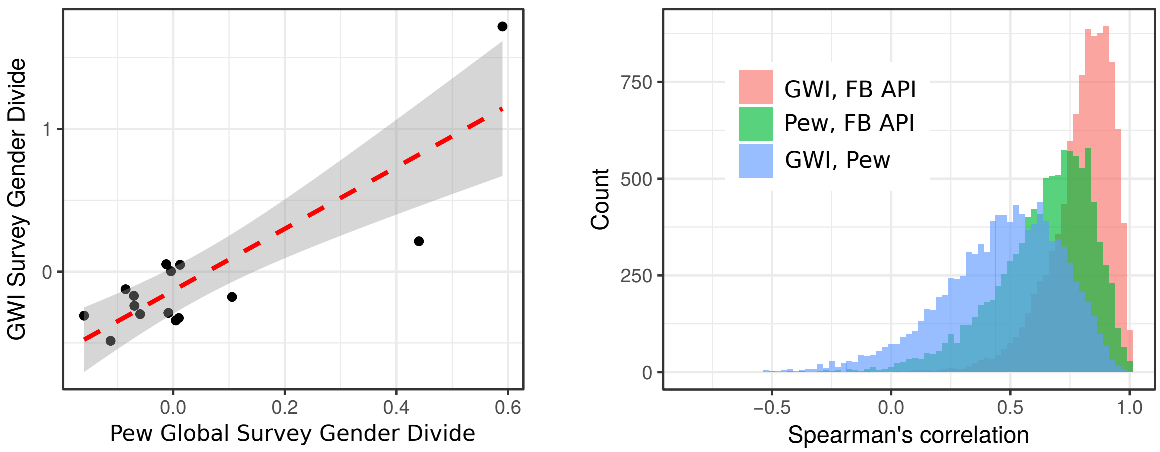

Figure 4: Left: Comparison of gender divide measurements in PEW and GWI data. The red dashed line shows a linear regression profile, with its prediction standard errors in the shaded area. Right: Bootstrapping distributions of Spearman's correlation coefficient between all pairs of gender divide measurements.

Spearman: 0.49, CI $[-0.11, 0.86]$). As with Facebook Penetration, the correlation between the FGD using the Facebook marketing API and the other two (with Pew: Pearson: 0.77, CI $[0.43, 0.92]$, Spearman: 0.68, CI $[0.14, 0.93]$; with GWI: Pearson 0.96, CI $[0.890.99]$, Spearman 0.83, CI $[0.47, 0.97]$) estimates is comparable to the correlation within estimates, as evidenced in bootstrapping samples reported in the right panel of Fig. 4. We can conclude that the estimate of the FGD using the marketing API is consistent with GWI and Pew survey metrics, opening the study of the FGD to a much larger sample of countries.

We measured the absolute difference between the FGD in the marketing API and in each survey dataset. As expected, the correlation between this absolute difference and the Facebook penetration across countries is negative (with GWI: Pearson: -0.3, CI $[-0.59, 0.05]$, p-value$= 0.09$; with Pew: Pearson: -0.27, CI $[-0.65, 0.21]$, p-value$= 0.26$), but its value is weak and not significant. Nevertheless, we include controls for Facebook penetration in our further analyses, to make sure that our results are not an artifact of a correlation between penetration and measurement error in the marketing API.

Furthermore, the GWI survey allows us to compare measurements of the FGD in other social networks with our measure based on the Facebook marketing API. We get moderate to high Pearson correlation coefficients with other sites, such as Whatsapp (0.67, CI $[0.43, 0.82]$), LinkedIn (0.65, CI $[0.40, 0.81]$), Twitter (0.69, CI $[0.46, 0.84]$), Instagram (0.79, CI $[0.62, 0.89]$), and YouTube (0.89, CI $[0.79, 0.94]$). While we cannot generalize to all social networks based only on Facebook data, we can see that, to some extent, the difference in activity across genders also appears in other SNS. This is particularly interesting when comparing Facebook, a very private social network, with YouTube or Twitter, which are much more public but still display substantial correlations in terms of FGD.

Comparison across age groups

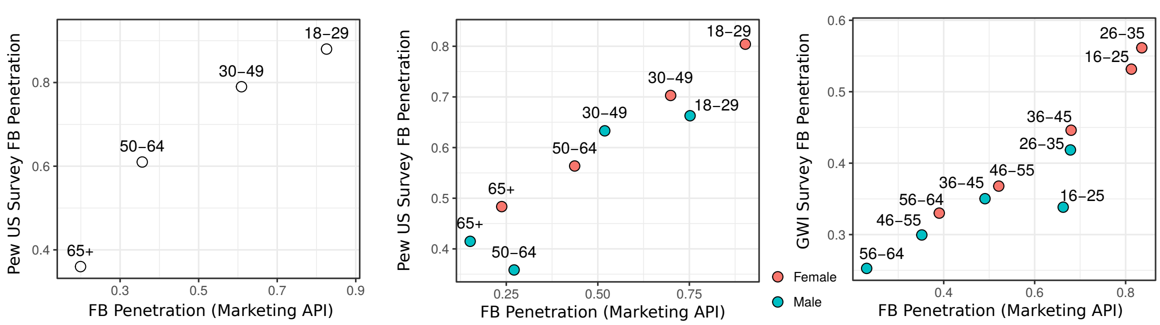

Figure 5: Comparison of FB presence ratio versus PEW gender categories for both genders together (left) and gender-wise (center). Replication of the same validation versus GWI estimates (right).

We compared Facebook penetration estimates in the US across the four age groups reported in the Pew US dataset. The left panel of Fig. 5 shows the comparison for both genders together, which have a Pearson correlation coefficient of 0.96, CI $[0.04, 0.99]$. The central panel of Fig. 5 shows the same comparison by taking the gender-wise estimates, which also have positive Pearson correlation (0.94, CI $[0.72, 0.99]$). This also appears when surveying the GWI dataset for US respondents in similar age categories, as shown on the left panel of Fig. 5, which has a high and significant Pearson correlation coefficient of 0.92, CI $[0.69, 0.98]$. We can conclude that data provided by the Facebook marketing API is consistent across ages, but to be sure that our further analyses are robust we take two action: 1) we add a mean user age control to our regression models, and 2) we stratify our analyses across age categories, using in each stratum a measurement of the FGD in the corresponding age range.

Subdaily measurement consistency

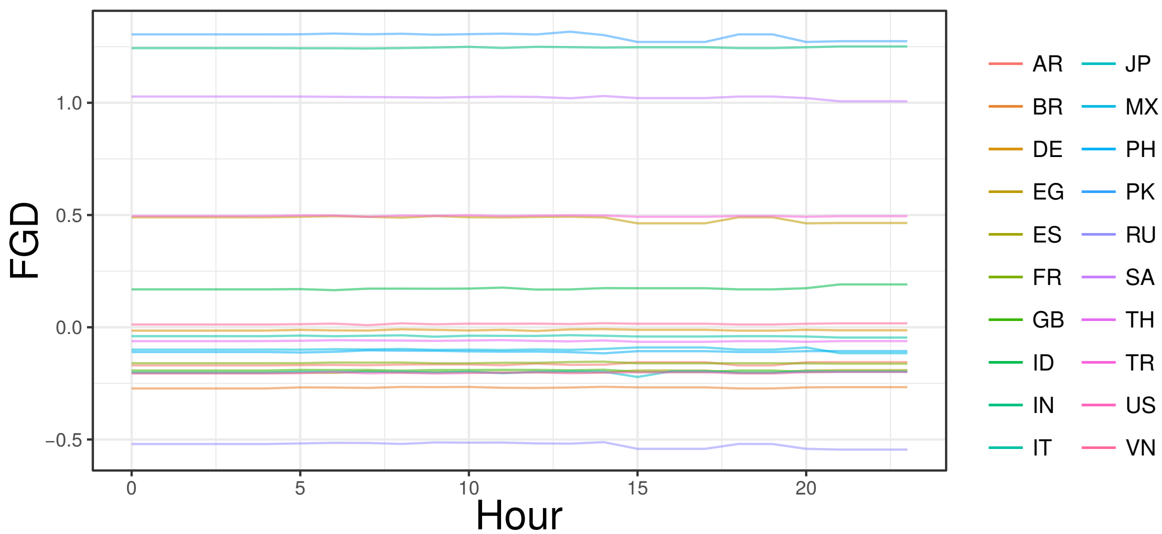

Figure 6: Examples of subdaily trajectories of the FGD.

We retrieved data from the Facebook marketing API on a daily frequency, starting our retrieval at 3 AM Central European Time. To validate the consistency of our measurement with any other times of the day, we checked the consistency of our construction of the FGD with hourly values for 24 hours in December 2017.

Fig. 6 shows the hourly measurement for a sample of large countries, revealing high consistency with very small fluctuations. When comparing the measurement of the FGD at 3AM CET with any other time in the same day, we get extremely high pearson correlation coefficients (0.9942654, CI $[0.9939277, 0.9945844]$), as also evidenced in Fig. 7. This also extends to the measurement of Daily Active Users for male (0.9999431, CI $[0.9999397, 0.9999462]$) and female (0.9999378, CI $[0.9999341, 0.9999412]$), as well as the total number of accounts for male users (0.9999926, CI $[0.9999921, 0.9999930]$) and female users (0.9999968, CI $[0.9999966, 0.9999970]$). Any fluctuation can be attributed to the rounding that Facebook does to preserve individual user anonymity and to the inter day changes in the number of Daily Active Users

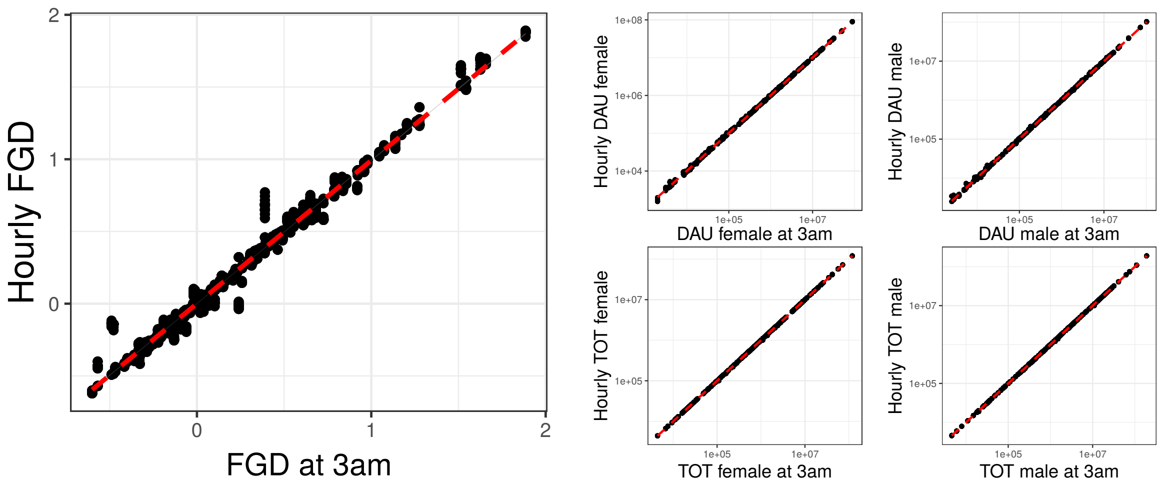

Figure 7: Facebook API measurements at different hours of the day. The left panel shows a comparison of measurements of the FGD for all countries in the dataset at 3AM CET versus hourly measurements at other times of the day. The right panels show the comparison between the measurement at 3AM CET and at other times of the day for the number of DAU and of present users (TOT) per gender.

Supplementary Text 2 - FGD as a function of other inequalities

Regression diagnostics

The Variance Inflation Factors of the variables in the FGD model are below 5, allowing us to discard collinearity in the linear model of FGD as a function of other inequalities. Table 2 reports the detailed results of the FGD model fit and Table 3 reports the results of the same model when fitted with a robust regression method. Table 4 shows a fit with HC correction for heteroskedasticity. All results are qualitatively similar, revealing that the FGD model result is robust to outliers and heteroskedasticity.

Term	Median estimate	95% Credible Interval	p-value
Intercept	$\mathbf{135.8}$	$[119.9, 152.1]$	$p < 0.01$
Education Equality Rank	$\mathbf{-0.54}$	$[-0.67, -0.41]$	$p < 0.01$
Health Equality Rank	$\mathbf{-0.27}$	$[-0.37, -0.17]$	$p < 0.01$
Economic Equality Rank	$\mathbf{-0.16}$	$[-0.27, -0.06]$	$p < 0.01$
Political Equality Rank	0.05	$[-0.05, 0.14]$	0.19
Internet Penetration Rank	$\mathbf{-0.27}$	$[-0.44, -0.09]$	$p < 0.01$
Income Inequality Rank	0.01	$[-0.09, 0.10]$	0.44
Population Rank	0.01	$[-0.09, 0.11]$	0.40
Facebook Penetration Rank	0.03	$[-0.12, 0.18]$	0.33
Mean User Age Rank	0.02	$[-0.08, 0.11]$	0.33
N	142	R^2	0.7417

Table 2: Regression results of FGD model. Estimates of p-values are based on the posterior of parameter estimates after 10,000 iterations.

Fig. 8 shows the normal Q-Q plot and the histogram of residuals, which are distributed very close to normality. This is confirmed by a Shapiro-Wilk normality test, with a statistic of 0.99 and unable to reject the null hypothesis that residuals are normally distributed ($p = 0.63$). Furthermore, residuals are uncorrelated with all gender equality variables (Near-zero Pearson correlation coefficients, with p-values above 0.9) and the square root of absolute residuals are not significantly correlated with predicted values. In addition, Facebook penetration is uncorrelated with residuals of the model (Pearson -0.0028, p-value $= 0.97$), showing no signs of bias due to the variance of Facebook penetration rates across countries.

Term	Estimate	Standard Error	t-value	p-value
Intercept	**137.4**	9.14	15.04	$p < 10^{-10}$
Education Equality Rank	$\mathbf{-0.55}$	0.07	-7.47	$p < 10^{-10}$
Health Equality Rank	$\mathbf{-0.29}$	0.05	-4.98	$p < 10^{-5}$
Economic Equality Rank	$\mathbf{-0.13}$	0.06	-2.03	0.04
Political Equality Rank	0.05	0.05	0.95	0.34
Internet Penetration Rank	$\mathbf{-0.29}$	0.10	-2.89	$p < 0.01$
Income Inequality Rank	0.001	0.05	0.01	0.99
Population Rank	-0.003	0.05	0.05	0.96
Facebook Penetration Rank	0.05	0.09	0.62	0.54
Mean User Age Rank	0.01	0.05	0.21	0.84
N	142	Multiple R^2		0.6715

Table 3: Robust regression results of FGD model.

Term	Estimate	95% HC CI	Standard error	p-value
Intercept	**136.8**	$[122.5, 151.2]$	7.32	$p < 10^{-10}$
Education Equality Rank	$\mathbf{-0.54}$	$[-0.68, -0.40]$	0.07	$p < 10^{-10}$
Health Equality Rank	$\mathbf{-0.27}$	$[-0.36, -0.18]$	0.05	$p < 10^{-8}$
Economic Equality Rank	$\mathbf{-0.17}$	$[-0.26, -0.07]$	0.05	$p < 0.001$
Political Equality Rank	0.04	$[-0.04, 0.13]$	0.04	0.32
Internet Penetration Rank	$\mathbf{-0.27}$	$[-0.44, -0.10]$	0.09	$p < 0.01$
Income Inequality Rank	0.004	$[-0.09, 0.10]$	0.05	0.92
Population Rank	0.01	$[-0.08, 0.10]$	0.05	0.87
Facebook Penetration Rank	0.03	$[-0.12, 0.18]$	0.08	0.7
Mean User Age Rank	0.02	$[-0.07, 0.10]$	0.04	0.66

Table 4: Coefficient estimates using HC corrected estimates.

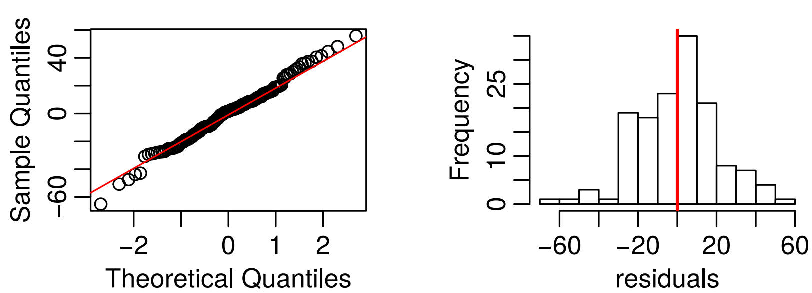

Figure 8: Left: Normal Q-Q plot of residuals of the FGD model. Right: Histogram of residuals.

The role of GDP

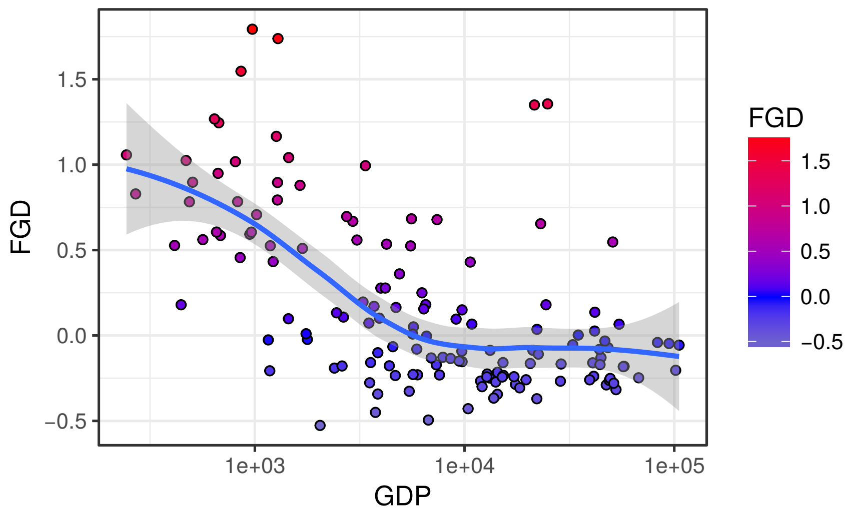

Figure 9: Relationship between FGD and GDP (log scale).

Fig. shows the relationship between the FGD and the GDP per capita. There is a significant negative correlation between both (Spearman -0.57, $p < 10^{-6}$), motivating a replication of the above model with GDP as a control. Since GDP is highly correlated with Internet Penetration and leads to high Variance Inflation, we replace Internet Penetration with GDP in our model. Coefficient estimates are reported on Fig. 10, revealing that the main result is robust to controlling for the wealth of countries.

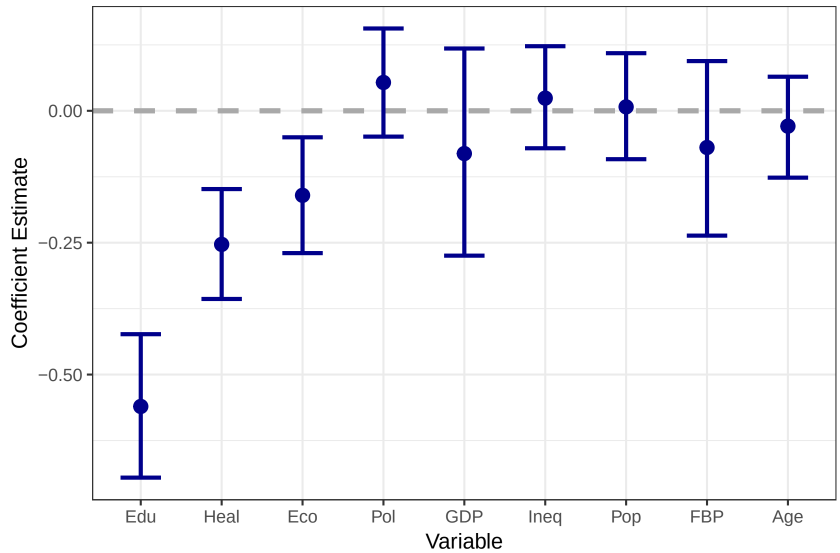

Figure 10: Coefficient estimates of the FGD model with GDP instead of Internet Penetration as control.

Model stability across monthly measurements

We repeated the fit of the FGD model for measurements of the DAU in twelve months between 2015 and 2016. Fig. 11 shows the results of the fit for these alternative periods. The coefficient estimates barely depend on the period when the DAU are calculated and the R^2 of the fits range between 0.727 and 0.758, confirming that our results are robust to fluctuations in the reporting of DAU through the Facebook API.

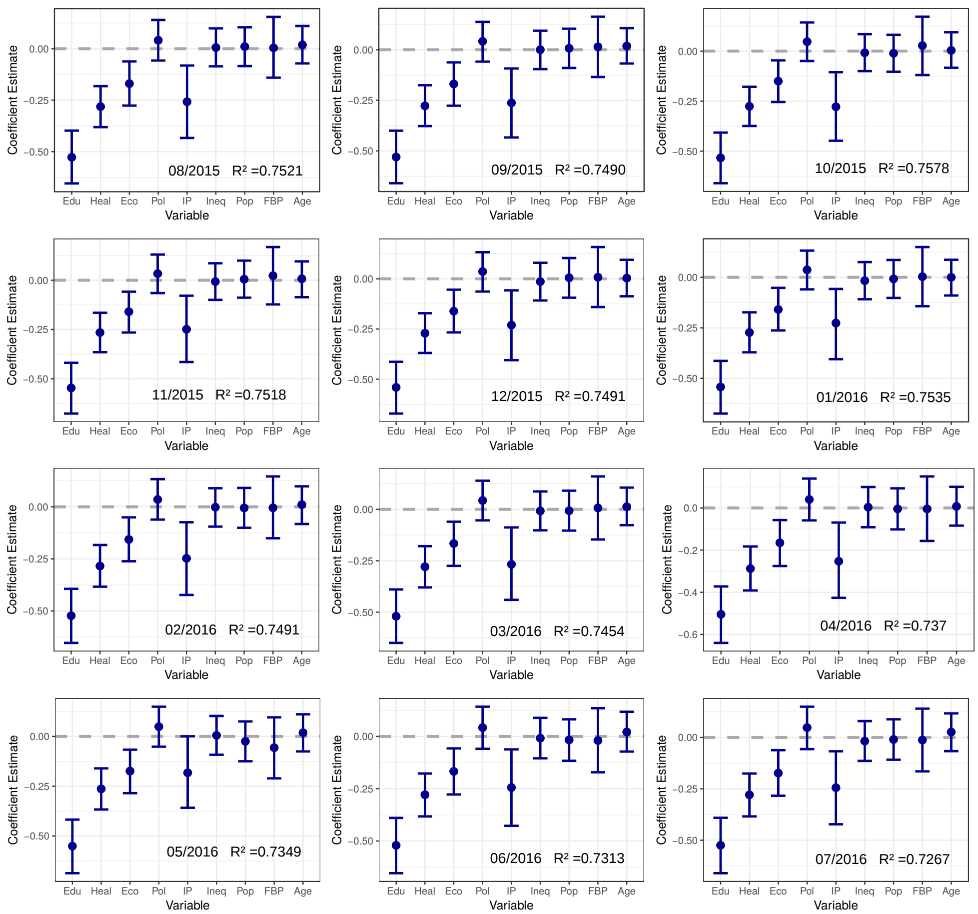

Figure 11: Results of repetition of the fit of the FGD model for 12 months between 2015 and 2016. The results are qualitatively stable across months.

Model stability across age groups

Figure 12 shows the replication of the model for segments of different age groups. Results are qualitatively similar to those of the whole population, with significant effects of gender equality variables and high R^2 values.

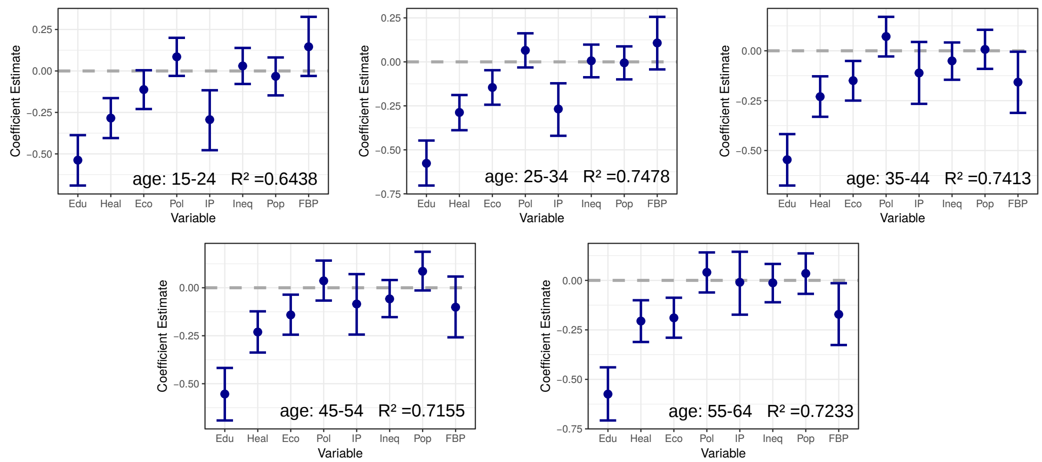

Figure 12: Replication of the model for data segmented into age groups.

Model test with GWI and Pew approximations to the FGD

We repeated the model using approximations of the FGD using the limited sample of GWI. The performance of the model is similar, as shown in Figure 13, with $R^2 = 0.77$. While the sample size of GWI is too small to test the role of all equality variables, the Pearson correlation between the rank of education equality and the rank of FGD in GWI is -0.52 ($p - value < 0.01$). Similarly, the model for the approximation of the FGD with data from Pew for all SNS gives similar $R^2 = 0.63$ and a significant negative Pearson correlation between the rank of education equality and the rank of FGD in PEW (-0.73, $p - value < 0.001$).

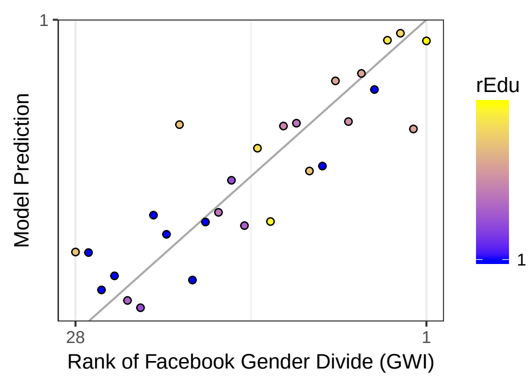

Figure 13: Replication of the model with the GWI estimate of the FGD.

Supplementary Text 3 - Network externalities

The results of the network externalities model are shown on Fig. 14. The model achieves a $R^2 = 0.96$ on the logarithmic scale and a $R^2 = 0.89$ on the linear scale of activity ratios per gender. Table 5 shows the detailed results of the model, evidencing the superlinear scaling ($\alpha = 1.2$) and the difference between genders ($\alpha_F = 0.25$).

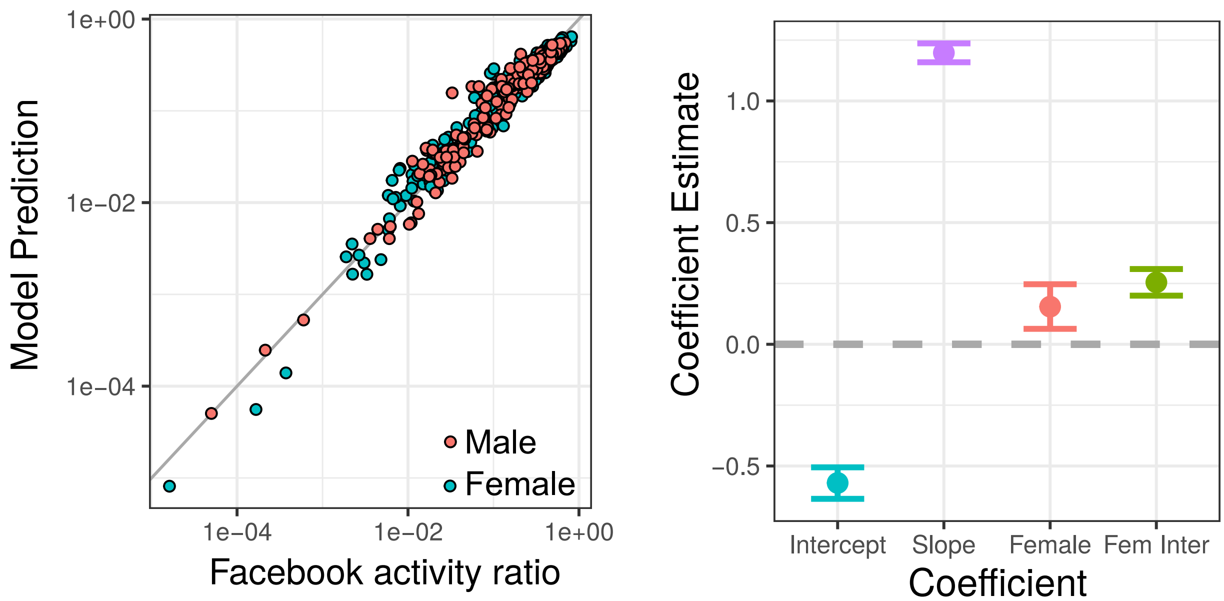

Figure 14: Results of network externalities model fit. The left panel shows the comparison between empirical and predicted values, the right panel shows median estimates and 95% CI of model coefficients.

Term	Median estimate	95% Credible Interval	p-value
β	$\mathbf{-0.57}$	$[-0.63, -0.50]$	$p < 0.01$
α	$\mathbf{1.198}$	$[1.16, 1.24]$	$p < 0.01$
β_F	$\mathbf{0.15}$	$[0.06, 0.25]$	$p < 0.01$
α_F	$\mathbf{0.25}$	$[0.2, 0.31]$	$p < 0.01$
N	422 (211 countries, 2 genders)	R^2	0.96

Table 5: Regression results of network externalities model. Estimates of p-values are based on the null hypothesis that the coefficient equals one for α and zero for the rest, after 10,000 iterations.

Fig. 15 shows the analysis of the residuals of the model and the error in the linear scale of activity ratios per gender. Some small deviations from normality can be observed at the tails, corresponding to significant Shapiro-Wilk statistics of 0.94 and 0.95. Both types of residuals are uncorrelated with Facebook penetration and do not appear to have a structure across predicted values. We identified some of the residual outliers, such as China and Tajikistan, which when removed do not have a qualitative impact in the results of the model fit and lead to residual distributions closer to normality.

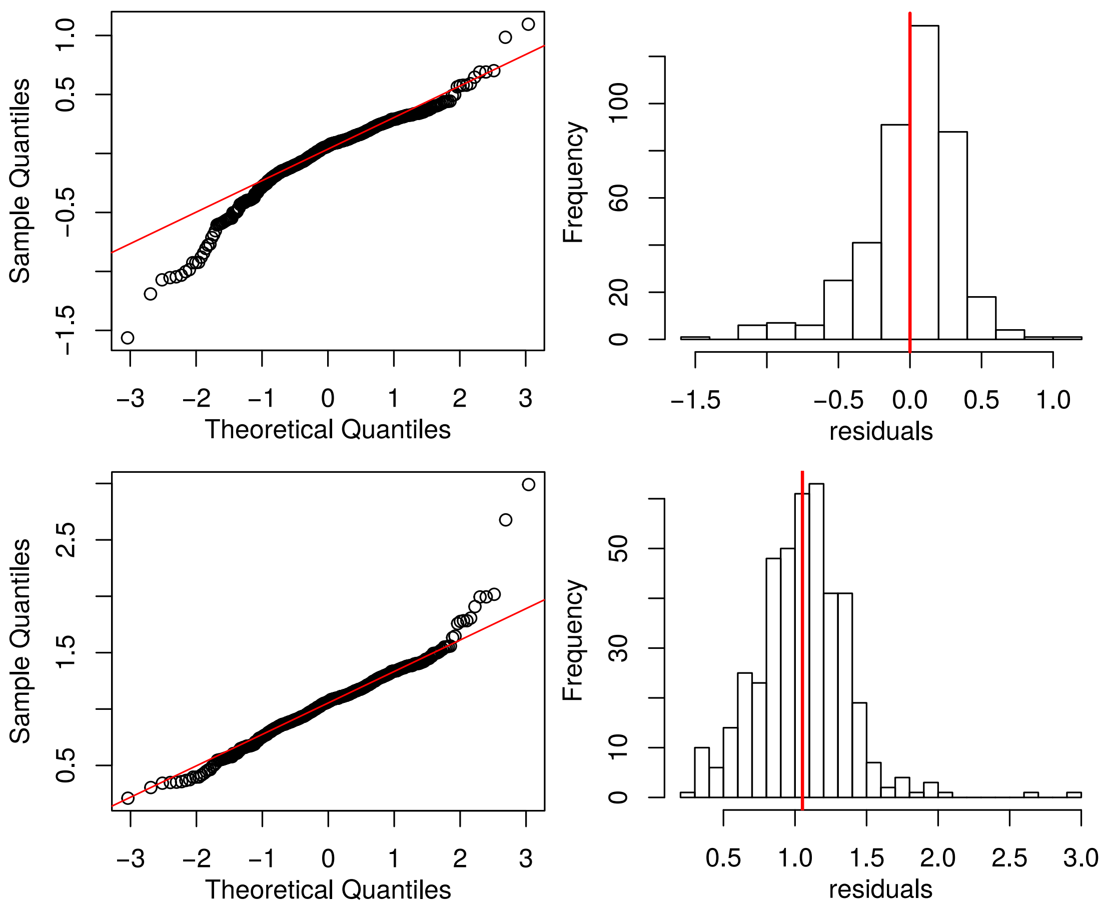

Figure 15: Analysis of residuals of network externalities model. The top panels show the normal Q-Q plot and the histogram of residuals ϕ of the model, and the lower panels the converse for multiplicative residuals in the linear scale of activity ratios per gender. Some minor deviations from normality can be observed in both.

Table 6 shows the results when correcting for heteroskedasticity. All results remain qualitatively unchanged.

Term	Estimate	95% Confidence Interval	Standard Error	p-value
β	$\mathbf{-0.57}$	$[-0.61, -0.53]$	0.02	$p < 10^{-10}$
α	$\mathbf{1.198}$	$[1.17, 1.23]$	0.02	$p < 10^{-10}$
β_F	$\mathbf{0.15}$	$[0.07, 0.24]$	0.045	$p < 0.001$
α_F	$\mathbf{0.25}$	$[0.18, 0.33]$	0.038	$p < 10^{-10}$

Table 6: HC corrected results of network externalities model.

We repeated the fit using a robust regression method, reporting the results on Table 7. While estimates slightly change, the qualitative results of a superlinear relationship that is stronger for female users still hold. This shows that our conclusions are robust to the influence of outliers.

Term	Estimate	Standard Error	t-value	p-value
β	$\mathbf{-0.523}$	0.033	-16.25	$p < 10^{-10}$
α	$\mathbf{1.19}$	0.02	60.88	$p < 10^{-10}$
β_F	$\mathbf{0.25}$	0.05	5.43	$p < 10^{-7}$
α_F	$\mathbf{0.37}$	0.03	12.59	$p < 10^{-10}$

Table 7: Robust regression results of the network externalities model.

As with previous models, we evaluated the model of network externalities over twelve months following our initial measurement. Fig. 16 reports the overall results, showing no relevant decrease in R^2 and generally the same result, where the parameter α_F is significantly larger than zero and the parameter α is significantly larger than one.

We stratified the analysis, fitting the network externalities model using calculations of the FGD using only data from a set of age categories. Fig. 17 shows the model results, evidencing that the female intercept, which measures the surplus of the exponent for female users, is positive and significant for all age categories.

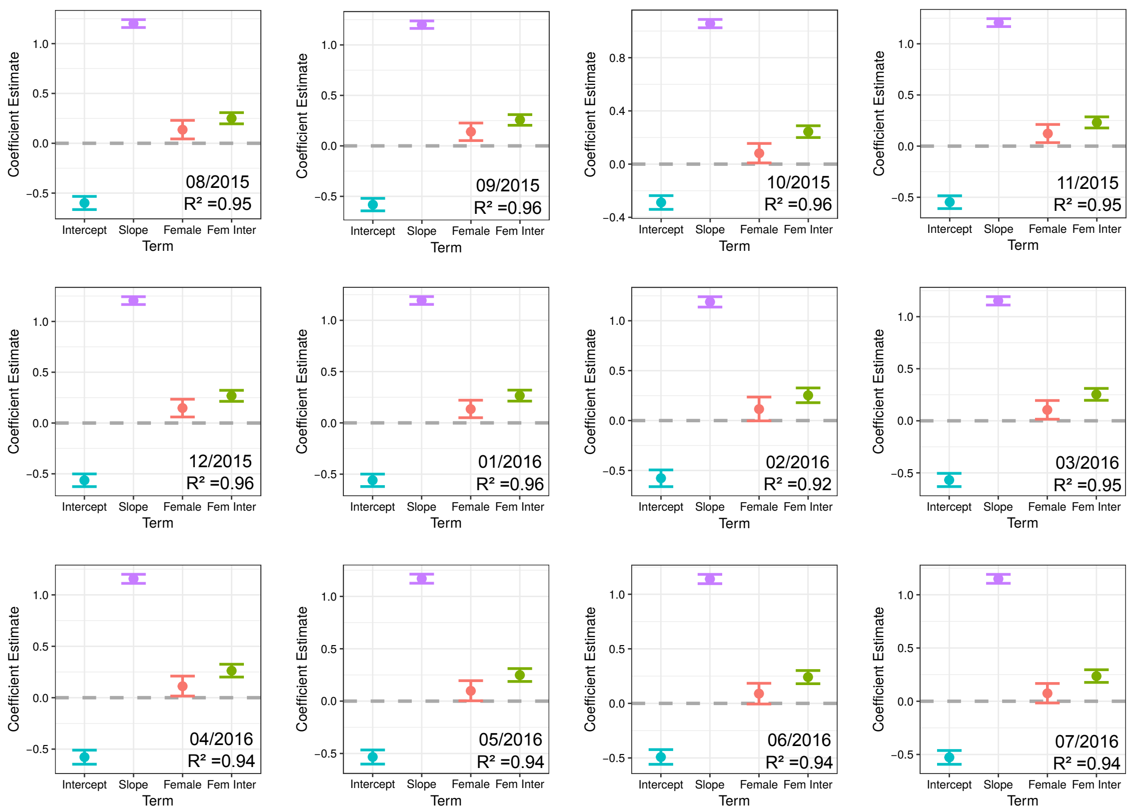

Figure 16: Results of repetition of the fit of the network externalities model for 12 months between 2015 and 2016. The results are qualitatively stable across months.

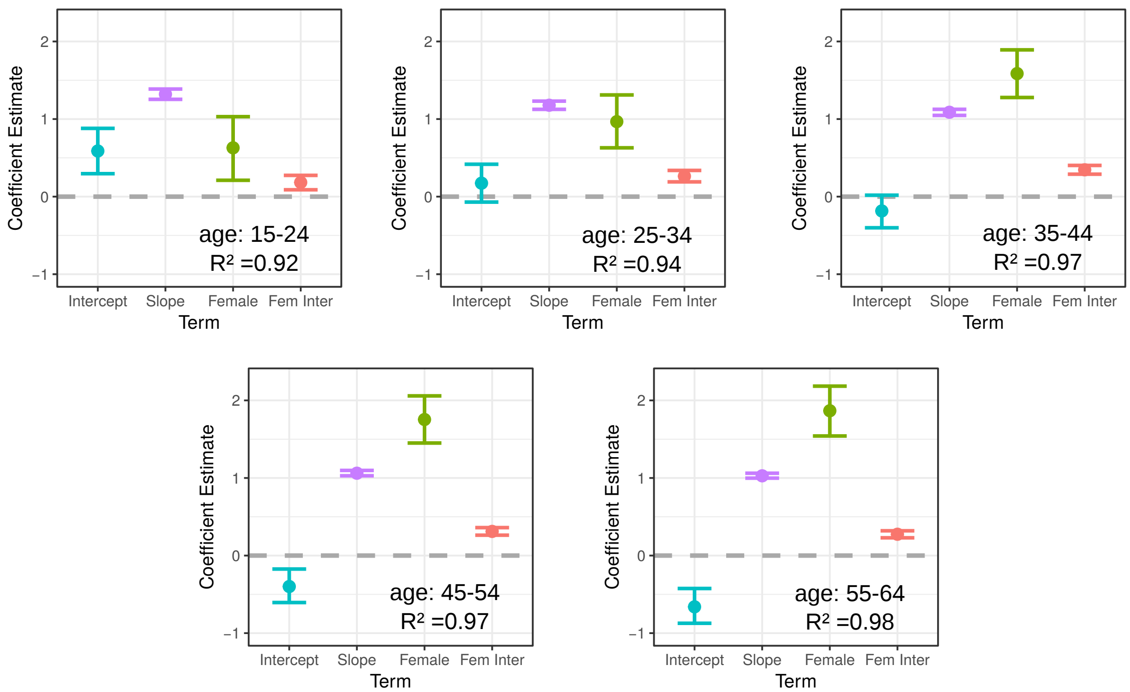

Figure 17: Results of repetition of the fit of the network effects model for age segments.

Supplementary Text 4 - Gender equality changes

Table 8 reports the Variance Inflation Factors of the variables in the model of economic gender equality changes and in the model of changes of FGD. All factors are low enough to discard multi-collinearity.

ΔEco Model		ΔFGD Model	
Variable	VIF	Variable	VIF
FGD Rank	1.833	FGD	1.66
Economic Gender Equality 2015	1.276	Rank Economic Gender Equality 2015	1.15
GDP Rank	1.505	GDP Rank	1.493

Table 8: Variance Inflation Factors of independent variables in the economic gender equality changes model.

Table 9 presents the detailed results of both models of changes. Before fitting, we rescaled the ranked variables to have a value between zero and one to allow a better comparison of their relationships, controlling for autocorrelation by including the unranked value of the variable in the previous year. The results of Table 9 are confirmed by ANOVA tests of the FGD rank in the ΔEco model $F = 10.195, p < 0.01$, and the non-significant result for the Eco rank in the ΔFGD model $F = 0.003, p > 0.9$.

ΔEco Model				ΔFGD Model			
Term	Estimate	s.e	p-value	Term	Estimate	S.e	p-value
Intercept	-0.011	0.015	0.437	Intercept	0.019	0.010	0.069
FGD Rank	**0.039**	0.01	< 0.01	FGD	-0.002	0.011	0.848
Eco	**-0.061**	0.021	< 0.005	Rank Eco	-0.001	0.015	0.94
GDP Rank	**0.052**	0.01	$< 10^{-6}$	GDP Rank	0.007	0.015	0.65
Multiple R^2		0.1501		Multiple R^2		0.0009	

Table 9: Results of robust regression for the model of changes in economic gender equality and of changes in the FGD.

The residuals of the model of changes in economic gender equality are distributed close to normality, as shown on Fig. 18, with a significant Shapiro-Wilk statistic of 0.97 and only some small deviations from normality at the tails. Residuals are uncorrelated with all independent variables and do not show signs of heteroskedasticity.

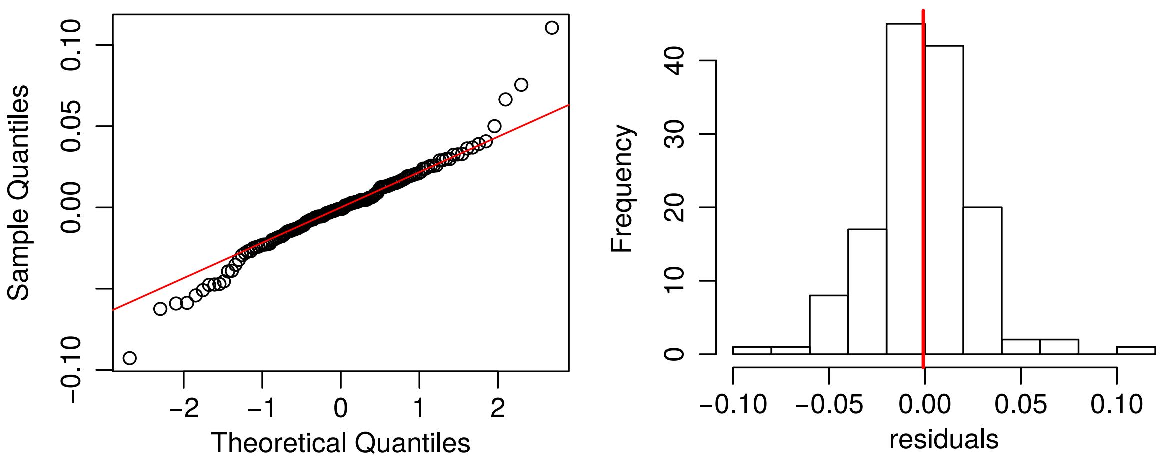

Figure 18: Analysis of residuals of the economic gender equality changes model. The left panel shows the normal Q-Q plot of residuals of the model, and the right panel their histogram. Some minor deviations from normality can be observed in both.

We tested the robustness of the positive association between FGD and ΔEco in two new fits including the same controls as for the FGD model (VIF of all factors below 5). The results are shown on Table 10, evidencing that the observed association between FGD and ΔEco is robust to other socio-economic indicators, and to other possible confounds such as Facebook penetration or mean user age.

We further tested the possible role of other equality indices in the relationship between FGD and ΔEco. We added all other three gender equality scores as controls (VIF below 5), and repeated the fit. The result, shown on Table 11 shows that the positive association between FGD and ΔEco is robust to the possible effect of other kinds of gender inequality.

Term	Estimate	s.e	p-value	Estimate	s.e	p-value
Intercept	0.007690	0.018282	0.67473	0.006294	0.017968	0.72667
FGD Rank	**0.024313**	0.010963	< 0.05	**0.026402**	0.010911	< 0.05
Eco	**-0.072524**	0.024797	< 0.01	**-0.075975**	0.024367	< 0.01
Ineq Rank	**-0.016436**	0.007985	< 0.05	-0.013190	0.007922	0.09833
Pop Rank	**0.017917**	0.008364	< 0.05	**0.019372**	0.008090	< 0.05
Mean Age Rank	0.005872	0.012132	0.62919	0.006279	0.011781	0.59494
FB Penetration Rank	0.015300	0.013533	0.26029	0.015453	0.012387	0.21442
GDP Rank	0.024484	0.016641	0.14359			
Internet Penetration Rank				0.024763	0.015389	0.11000
Multiple R^2	0.2009			0.1966		

Table 10: Results of the ΔEco model including additional controls.

Term	Estimate	s.e	p-value
Intercept	-0.029304	0.021736	0.179917
FGD Rank	**0.039560**	0.016828	< 0.05
Economic Score	-0.043693	0.025742	0.091983
GDP rank	**0.045928**	0.012187	< 0.001
Education Score rank	0.001291	0.013318	0.922924
Political Score rank	0.014899	0.009621	0.123862
Health Score rank	0.004209	0.009031	0.641953
N	139	Multiple R^2	0.171

Table 11: Results of the ΔEco model including controls for other gender equality indices.

We tested whether the association between FGD and ΔEco could be explained by general cultural differences. We combined our dataset with Hofstede's cultural dimensions: Power Distance Index (PDI), Individualism (IDV), Uncertainty Avoidance Index (UAI), and Masculinity (MAS) (VIF below 5). This limits the analysis to a set of 66 countries common to both datasets, with results reported on Table 12. The positive association between FGD and ΔEco is still significant, suggesting that the relationship between both variables goes beyond what Hofstede's model captures in terms of culture.

Following the same methodology as for previous models, we stratified our analysis with FGD measured only in a variety of age groups. The results are shown on Fig. 19, revealing the positive and significant role of FGD in the model for all age segments.

Term	Estimate	s.e	p-value
Intercept	0.04743	0.03102	0.1316
FGD Rank	**0.03792**	0.01764	< 0.05
Eco	**-0.1243**	0.03779	< 0.01
PDI	$3.217*10^{-4}$	$1.759*10^{-4}$	0.0725
IDV	$1.754*10^{-4}$	$-1.708*10^{-4}$	0.3088
MAS	$-2.490*10^{-4}$	$1.546*10^{-4}$	0.1126
UAI	$3.148*10^{-5}$	$1.366*10^{-4}$	0.8185
N	66	Multiple R^2	0.362

Table 12: Results of the ΔEco model including controls for cultural dimensions.

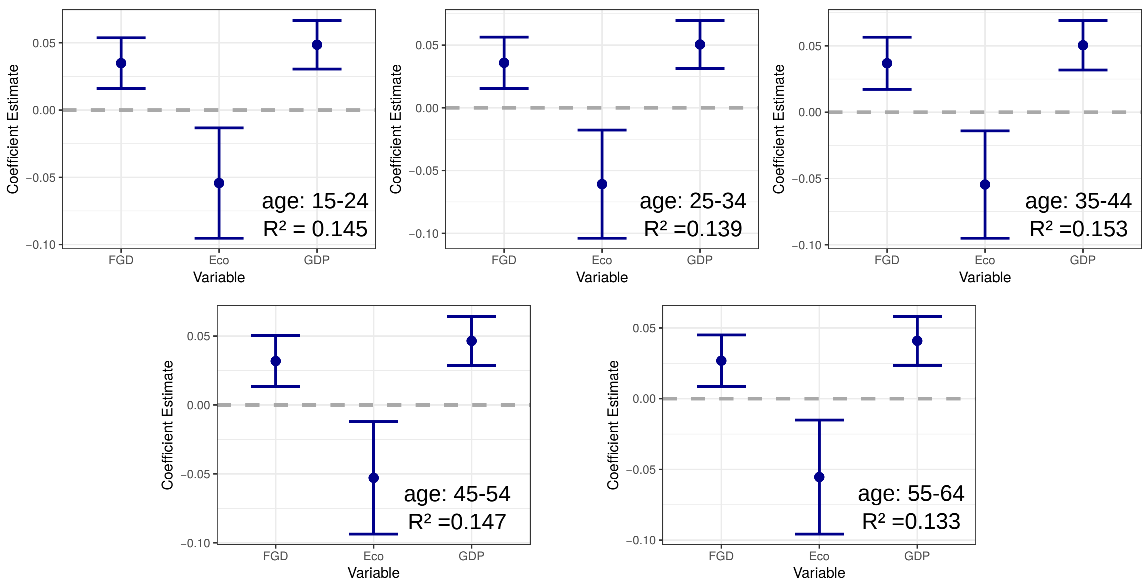

Figure 19: Replication of the model stratifying for different age groups.

Our data offers the opportunity to measure the role of FGD in the changes of other gender equality measures, but the indices for Political and Health gender equality have negligible changes between 2015 and 2016. For that reason we can only evaluate the role of FGD for changes in Education gender equality. We find no significant effect of FGD, as reported on Table 13.

Term	Estimate	s.e	p-value
Intercept	**0.0613210**	0.0613210	$< 10^{-7}$
FGD rank	-0.0004546	0.0023720	0.848
Edu	-0.0606702	0.0099782	$< 10^{-6}$
GDP rank	**-0.0012934**	0.0022238	0.562 < 0.01
N	139	Multiple R^2	0.08

Table 13: Results of the ΔEdu model.