

Inequality in Posting Behaviour Over Time

A study of Danish Facebook users

Anja Bechmann

Aarhus Institute of Advanced Studies, Aarhus University, Denmark, anjabechmann@aias.au.dk

Abstract

This study investigates the Facebook posting behaviour of 922 posting users over a time span of seven years (from 2007 to 2014), using an innovative combination of survey data and private profile feed post counts obtained through the Facebook Application Programming Interface (API) prior to the changes in 2015. A digital inequality lens is applied to study the effect of socio-demographic characteristics as well as time on posting behaviour. The findings indicate differences, for example in terms of gender and age, but some of this inequality is becoming smaller over time. The data set also shows inequality in the poster ratio in different age groups. Across all the demographic groups, the results show an increase in posting frequency in the time period observed, and limited evidence is found that young age groups have posted less on Facebook in more recent years.

Keywords: inequality, demographics, participation, Facebook, big data

Introduction

Posting on Facebook, as the world's largest communication platform, provides people with a voice in the digital socio-cultural and civic debate. The framing and agenda-setting theories (e.g. Scheufele, 1999) have shown how important it is for communication power to be able to set the frame in a discussion. At the same time, Facebook data have been available for third parties to use for internet profiling (Bechmann, 2014, 2015; Bruns et al., 2018). Having a voice when posting has therefore become a dual-edged sword that also potentially means privacy violation by exposing opinions and intimate and personal information to data mining by third parties. Self-censorship can be enacted not because people do not want to voice their concerns, take control in discussions or gain social power through self-portraits but because they want to protect their privacy (Baumer et al., 2013; Ertiö et al., 2018). Despite the positive consequences for individuals' communication power and the negative privacy-related consequences of posting on Facebook, few studies have tried to provide a longitudinal study of potential differences

Bechmann, A. (2019). Inequality in posting behaviour over time: A study of Danish Facebook users. *Nordicom Review*, 40 (Special Issue 1): 31-49. doi:10.2478/nor-2019-0012.

in such Facebook posting patterns within the field of media and communication studies. We simply know very little about whether some socio-demographic groups post significantly more on Facebook than others over time and whether large gaps exist between posting and non-posting users over time. Scrutinizing how the Facebook posting frequency and ratio develop over time and in relation to socio-demographic variables will provide a stronger empirical background for discussing issues (power and privacy) of potential inequality in our social media-informed data-driven society.

Criticism of communication data traces and APIs as a data source for research, societal predictions and subsequent actions is manifold but especially concerns the use of Facebook data as a “god’s eye view” on usage and users behind the digital traces (Blank, 2017; Blank & Lutz, 2017; Hargittai, 2015; Kitchin, 2014: 167). There is particular concern about the dominating data scientist approach whereby, as long as there are enough data on enough users, it will be possible to detect “true” patterns that can subsequently be applied to the greater population (Bechmann, 2019; Bowker, 2014; Schradie, 2013). The argument is that this approach results in a research design that is biased by potential participation inequality – not only in the form of selection bias by studying only active posting, commenting, sharing and liking users (Lomborg & Bechmann, 2014) but also by the way in which socio-demographic profiles use Facebook differently, for instance in terms of how much they post on Facebook profile feeds over time (Hargittai, 2015).

The aim of this study is to provide an empirical trace data-driven analysis of potential differences in Facebook posting patterns over time. The data set consists of survey demographic data in combination with API data from 1,000 Danes’ private Facebook profile feeds collected to mirror the Danish Facebook population on demographics. The data were collected for a seven-year period from when they joined Facebook until August 2014, before Facebook closed down its API access (Bruns et al., 2018). The article will examine whether this study found significant developments in posting frequency over time in general and differences in posting behaviour over time between various demographic variables in the data set (RQ1) and the share of posting users in different age groups over time (RQ2). The findings will feed into the discussion on the use of big (social media) data in a national population and in particular on the role of inequality as it plays out between frame setting as communication power and self-censorship as personal privacy protection.

Differences in posting behaviour on Facebook

Even though specific knowledge of inequality in posting behaviour developments on Facebook over time is scarce, related survey studies on digital divides and inequality in online and social media participation and between different social media user roles have provided interesting results on self-reported inequality in social media activities (e.g. Correa, 2010; Ertiö et al., 2018; Hargittai & Walejko, 2008; Nonnecke & Preece, 2000; Schradie, 2011).

The article builds on the claim (Blank & Lutz, 2017; van Dijk, 2006) that there is a particular need for more longitudinal studies to understand participation (here posting) inequality developments over time according to user demographics. The development over *time* is absent in most data-driven and survey studies on posting behaviour and other types of online and social media activities, despite this identified need. Some stud-

ies have tried to address time as a variable in connection with social media activities and sociodemographic inequality. For instance, when comparing Pew reports, Schradie (2011) found that the number of participating users on average rose from 5 per cent in 2005 to 21 per cent in 2008. However, the definition of participation broken down on social media networks is very imprecise. The overall framework is digital participation as creative content, and it is unclear exactly what that includes on social networks. Hargittai (2007) also suggested time spent on the internet as a predictor of participation, indirectly referring to the longitudinal perspective. The article thus builds on the assumption that the study will find an overall increase in posting during the period studied (Hypothesis 1).

Previous studies have shown that *education* is an important variable to consider when measuring inequality online. Certain (elitist) socioeconomic profiles participate more than others (Correa, 2010; Hargittai & Walejko, 2008; Schradie, 2011). Hargittai and Walejko (2008) conducted an important survey-based study of first-year college students in Chicago, showing that creative activity (defined as the creation of content and its sharing; OCC) on digital media in general is not randomly distributed among a diverse group of young adults. Rather, it is related to the socioeconomic status of the users (and their parents), favouring users with parents with a graduate degree as the most contributing users. They stressed the importance of examining such participation divides in future research because of the potential to create “social inequality” as “online content becomes increasingly important in setting social, political and cultural agendas” (Hargittai & Walejko, 2008: 253). Continuing this study, Schradie (2011) employed representative survey data from Pew internet research (US) to examine the socioeconomic inequality in online production outlined, for instance, on social networks. She found that education was the most robust predictor of inequality in participation. In his review of studies on digital inequality in general, Lutz (2016) discovered that education/income (socioeconomic status), age and gender were the most common indicators of inequality, and Ertiö, Kukkonen and Räsänen (2018) found these predictors to be significant for the differences in the types of social media activities that users report (sharing, commenting and posting). However, Blank (2013) concluded that education may be a strong predictor of general internet use but not of social media use (Blank & Lutz, 2017). Given the lack of data-driven studies, the article builds on the assumption that, despite education not being a strong predictor of the use of social media (or not), it is a significant predictor of posting frequency over time and will show unequal posting frequency over time in which highly educated people will post more than lower-educated people in the data set (Hypothesis 2).

Most studies have found *gender* to be a strong predictor of participation inequality. Trying to address the lack of research on how demographics influence specifically Facebook activity, McAndrew and Jeong (2012) found in their cross-national online survey (n=1,026) with a broad age group of 18-79 (mean age 30.24) that women engage more in Facebook activity than men. Activity is not understood as uploading content but rather more broadly as both spending time on Facebook and looking at or posting content. This finding is in line with a Finnish survey study (Ertiö et al., 2018) in which gender was shown to have an effect on differences in social media activities, in which women are slightly more active than men. The article therefore assumes that gender will have a significant effect on the posting frequency on Facebook over time and that

women will post more than men (Hypothesis 3). However, Lutz, Hoffmann and Mecke (2014) accounted for differences in inequality depending on the forms of participation, domains and topics studied (e.g. politics versus culture).

All the existing studies have found *age* to be a strong predictor of differences in posting patterns, especially content-related ones, on social media (Blank & Lutz, 2017; Ertiö et al., 2018; Hargittai, 2015; Lutz, 2016; McAndrew & Jeong, 2012). Younger users are more likely to use participatory media (Blank, 2013; Blank & Lutz, 2017; Correa, 2010; Deursen & Dijk, 2015; Hargittai & Walejko, 2008; Hoffmann et al., 2015). McAndrew and Jeong (2012) showed that older people spend less time on Facebook, have fewer friends and perform less of all the activities (also posting) than younger age groups. This finding was supported by Ertiö, Kukkonen and Räsänen (2018) finding that the older the user, the less likely he or she is to engage in social media activities of any kind. Even though these studies employ a different method (survey) for different samples, the study expects age to be a strong, significant predictor of posting frequency over time and to show that younger people post more frequently than older people over time (Hypothesis 4).

The study also considers the potential role of *area of residence*. None of the existing studies have included area of residence as a variable to test in relation to inequality in posting frequency. However, the variable is included primarily as a sample strata variable, secondarily as an analytical variable in this study. Due to the lack of existing studies, the study cautiously assumes that the area of residence will have no significant effect on the posting frequency measured over time and will show no unequal posting behaviour between urban and rural participants in the data set when (as in the other cases) adjusted for the effects of other (presumably correlating) variables, such as age and education (Hypothesis 5).

Since the early days of internet communities, studies have had an interest in examining the inequality between *social media user activity roles* (Bechmann & Lomborg, 2013; Hargittai, 2007; Nonnecke & Preece, 2000). Nonnecke and Preece (2000) were particularly interested in trying to outline “lurkers”. They discussed the difficulty of making a distinction between posting and lurking and how we conceptually should draw the line between the two. If someone posts in a certain period and not in others, or if they post once and never again, should they count as lurkers? They chose to define lurkers as users who have “either no posts or some minimal number of posts over a period of time” (Nonnecke & Preece, 2000: 2), finding that the percentage of lurkers or non-posters varied greatly for different discussion lists, ranging from one per cent to 99 per cent of the total number of users. Facebook-specific behavioural studies have been sparse on this topic, but survey studies have addressed the issue. Young (2011) reported, from a survey study of Australians between 15 and 65 years old, that 40 per cent of the respondents posted, but these studies did not measure the participation over a longer time period. This result is almost mirrored in a survey study from Finland on general social media activity roles in which 41.6 per cent posted original content online (Ertiö et al., 2018). Age was again by far the strongest predictor of social media posting behaviour, younger people being significantly more active in all posting behaviour than older people (Ertiö et al., 2018). As age is the strongest demographic variable in the existing studies, the study expects to find that around 40 per cent of people in the data set are posters over the whole period but that the percentage of posters is higher in the younger age groups than in the older ones (Hypothesis 6).

Method

Data collection

To answer the research questions and hypotheses, data were collected in the period May-July 2014 from a data set consisting of a total of 1,176,322 profile feed posts (N=1,000) from Danes with a Facebook account. The private data were collected with the permission of the participants and the Danish Data Agency using the Digital Footprints software (Bechmann & Vahlstrup, 2015), which facilitates data collection through the API. In May 2014, 67 per cent of the Danish internet population aged between 16 and 89 years had used social networking services. Out of these, 95 per cent had a Facebook profile, making Facebook the most popular social medium in Denmark, as in the rest of the Western world (Wijas-Jensen, 2014). With an already-high penetration rate of 67 per cent in 2014 (*ibid.*), Denmark is a good probe for testing posting frequency decreases over time and sociodemographic parameters. Whereas API data studies are limited to the information structure provided by Facebook (Lomborg & Bechmann, 2014), API data studies avoid interviewer effects, lack of memory (Junco, 2013) and other sources of measurement errors connected to survey methods (Lewis et al., 2008).

Representation

To overcome the representation issues of big data (Bowker, 2014), the study crosses borders between recruitment techniques in traditional surveys and APIs as a method for data collection (Bechmann & Vahlstrup, 2015). In contrast to many server data studies conducted in collaboration with Facebook (Bernstein et al., 2013; Burke et al., 2010), the data set is made up of participants recruited outside Facebook and validated externally through the online non-probability panel Userneeds. The widely used convenience sampling method consists of collecting data on Facebook through viral messages or Facebook ads (thereby obtaining many participants and a large data pool). This method is problematic, because studies have shown that people deliberately indicate false demographics to avoid certain ads and for privacy reasons (Bechmann, 2015; Marwick & Boyd, 2014). That is the reason for wanting to avoid this method to answer the research questions.

Out of 150,000 users, a largely representative sample of the Danish population, 2,898 users (with a Facebook account) stratified on demographics were invited to participate, for whom the Digital Footprints data transfer failed in 16 instances due to a change of password (3) and deletion of the token (13). A total of 1,000 users participated successfully in our study (34.5%). The low percentage was expected, as this is not a normal survey but rather an exchange of private data; this may result in sample measurement errors that will be discussed later in the article. The 1,000 participants were between 15 and 91 years old at the point of data collection, and 96 per cent used their own name on the Facebook profiles that were collected in the data set. No official Danish figures on detailed demographics exist, but we chose to use population estimates from the national internet panel for Denmark (Gemius) to stratify the data collection.¹

Opt-in online services are not ideal for estimating population values or claiming representative research as such, not least in this type of study, with private data collection and low response rates. The study focuses on describing the patterns in the data set (Baker et al., 2010). As the study has tried to mirror the Danish Facebook population

in the sample, we also find it informative to consider how the data map against this population despite the challenge of population estimates and collection methods. The study investigates whether significant results are obtained in the level of posts in general and combined with a given demographic parameter. The analysis measures significant tendencies by using the strata in the sample and discusses the differences observed in the two estimates.

Sociodemographics

Most studies of Facebook inequality have used a convenience sample, often of college students (Correa, 2010; Hargittai & Walejko, 2008). In such cases, it is not possible to extrapolate conclusions from patterns identified in these studies to the broader (national) population, because college students (especially regarding their age and education) are not representative of the total population on Facebook. Instead, we use a broad stratified sample to mirror the national population (Schradié, 2011) on the sociodemographic parameters of gender, age, education and area of residence,² as shown in Table 1 (see next page).

Instead of Facebook's own numbers, the study uses the estimated Facebook population of the official national internet traffic agency Gemius to stratify the sample. Table 1 shows that the sample has a slightly older and more urban profile than the Gemius estimated population. Furthermore, people with a short education are far less represented in the data set than in the general Gemius Facebook population. Even though the panel contains fewer participants with a short education than the population, the numbers are so different that they could point to a problem in the formulation of the survey question. The survey allowed the participants to indicate whether they had a short, medium or long education, explaining what these levels meant. However, it did not ask for specific education. This might have resulted in participants falsely registering a higher educational background. In any event, the article includes education but cannot show significant developments due to the small sample size for this variable. Despite this measurement error, the sample allows for a more substantial approach, examining patterns across age instead of relying on a data set with a homogeneous group of, for instance, students (like many existing trace data studies, e.g. Lewis et al., 2008).

Time and longitudinal study design

The data collection ended in July 2014, and the access token allowed data access for 60 days. This means that the newest data are from August 2014, but May-August lack some of the data in the data set. To adjust for these differences, the article uses data from March 2014 and backwards to achieve a balanced data set. Furthermore, we can see from the data that the first post in the data set appeared in 2007. Therefore, the article analyses the development from 2007 to March 2014.

The ability to collect data backwards in the Facebook API provides a potential detailed account of the way in which the amount of posts varies during the years. However, the time when the participants joined Facebook varies, resulting in a smaller number of people at the beginning than at the end of the period. Therefore, the analysis measures the statistical significance of the tendencies to adjust for these differences.

Table 1. The data set compared with the Gemius estimated total population of Facebook users in Denmark

	Digital Footprints data set (August 2014)		Gemius Facebook user data set (August 2014)		Facebook users in DK (Gemius, August 2014)		Danish population (August 2014)		
Age 15-29	239	23.9	616	15.5	926,384	29.1	-5.2	1,042,898	24.7
Age 30-44	191	19.1	835	21.0	797,608	25.1	-6	1,013,786	24.0
Age 45-59	282	28.2	1,428	35.8	827,729	26.0	2.2	1,153,027	27.3
Age 60+	201	20.1	1,108	27.8	628,619	19.8	0.3	1,010,763	23.9
Female	518	51.8	1,858	46.6	1,604,057	50.4	1.4	2,052,919	48.6
Male	482	48.2	2,129	53.4	1,576,282	49.6	-1.4	2,167,555	51.4
Short education	88	8.8	2,172	54.5	1,793,297	56.4	-47.6	2,282,398	54.1
High school, colleges and equivalent educa- tion	381	38.1	1,157	29.0	884,085	27.8	10.3	1,202,341	28.5
Higher education	531	53.1	658	16.5	502,958	15.8	37.3	735,735	17.4
Rural	543	54.3	3,022	75.8	232,1514	73.0	-18.7	3,146,063	74.5
Urban	457	45.7	965	24.2	858,826	27.0	18.7	1,074,411	25.5
Total	1,000	100	3,984	100	3,180,340	100	-	4,220,474	100

Definition of posters and posts

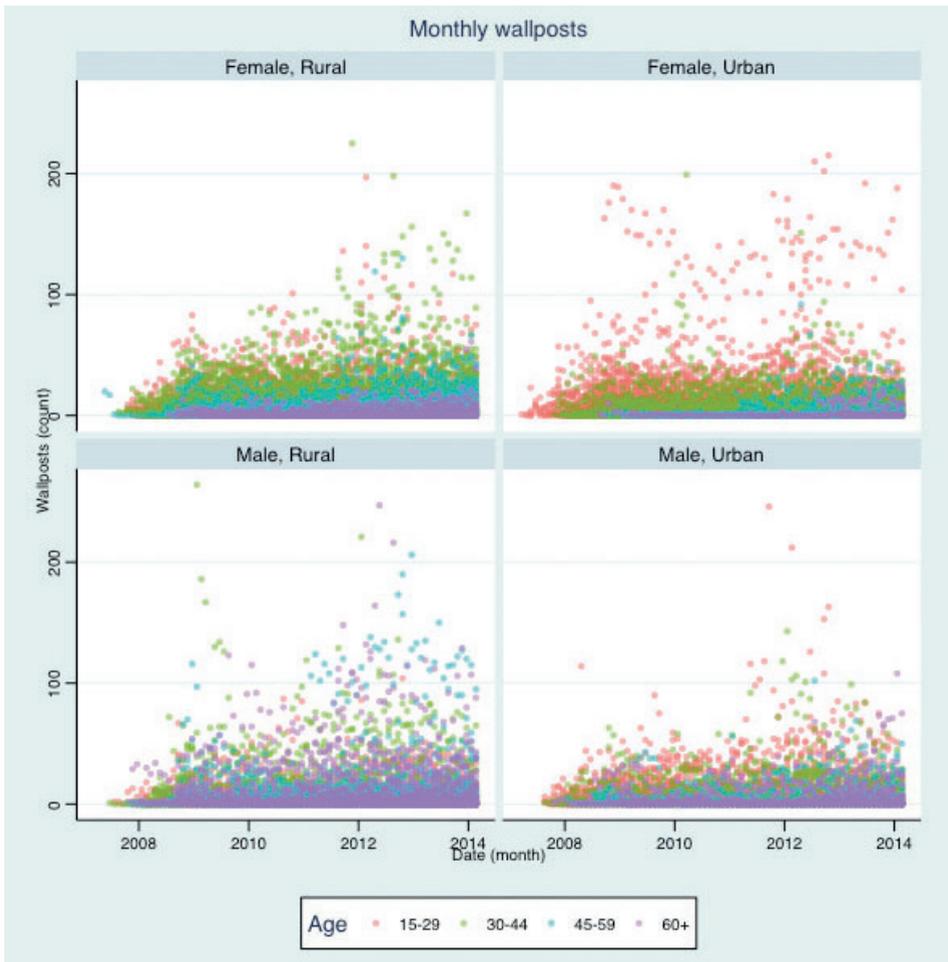
Inspired by Nonnecke and Preece (2000), this study considers a “poster” to be a user who posts at least once per month and a non-poster (with a Facebook account) to be one who posts less than once a month (including not at all). The data set only shows people joining, not leaving, and the study does not account for people deleting posts, only the remaining posts. In their Facebook data study, Bernstein and colleagues (2013) defined posts as status updates and link shares. However, since 2013, image and video uploads have become common on Facebook and are therefore included in the definition of a post in this study. The study focuses on posts and not comments, likes and group, page and Messenger postings, because posts set the frame for discussions similar to what existing studies have defined as own/creative content (Ertio et al., 2018; Hargittai & Walejko, 2008; Nonnecke & Preece, 2000). Groups and pages have not been consistent features in the period studied. Some of the features in the interface of Facebook and the infrastructure of the API moreover may have changed over time (Bechmann & Vahlstrup, 2015). Overall, the study does not take into consideration such changes in the interface and the structure of the API. This means that there might be a measurement error at the beginning of the period in an earlier version of the API, leading to fewer posts being transferred. Furthermore, server breakdowns or other technical errors might result in fewer posts being transferred to the Digital Footprints database (ibid.). However, to adjust for this, we implement alert functioning, indicating when data were not transferred, and manually update the transfer. Looking at the posting patterns, we find large fluctuations, as shown in Figure 1 on the next page.

As illustrated in Figure 1, the data set shows a particularly large peak in the posts in 2012, but this peak was caused by a small number of users. The outliers are auto-generated posts from apps that were allowed to post on behalf of users (acting as the user). The app Farmville was a particularly aggressive “poster”, for instance in one case having a monthly posting rate of 7,000 posts. Even though we only focus on users’ own posts and not apps, these types of posts were still part of the data set. The participation activity is interesting in itself, but, according to our definition of posts, we needed to remove them from the data set. If we blacklisted all apps from the data set, we would also delete all the posts made from Facebook’s native apps. Instead of blacklisting apps, we whitelisted all Facebook native apps (e.g. 47,151 came from Facebook for iPhone, 29,863 were made on Facebook for Android and 9,233 were from Facebook for iPad). In total, 17 Facebook apps were whitelisted. This provided us with a mean of 6.34 posts per month, a median of 2 posts and a standard deviation of 11.38.

Analytical method

As the data set period was filtered to March 2014, the filtered data set is N=960. This means that 40 participants joined after March 2014. When adjusted for the apps posting on behalf of the users, we ended up with N=922.

To account for posting developments, we choose to report the Poisson regression models primarily instead of the negative binomial models, because the Poisson models provided more conservative results. The negative binomial model in one case created a different trend that is briefly reported in the result section. The Poisson model is often used for count data (such as profile posts) specifically characterized by taking only non-

Figure 1. Monthly profile posts before normalization

negative integer values and, as the case is here, with a relatively low posting frequency, being non-normally distributed (Cameron & Trivedi, 2010). In particular, a random-effect estimator is used to account for the posting trends across gender, age (dynamically age adjusted), education and residence in separate models. Despite Durbin–Wu–Hausman tests indicating inconsistency between the fixed-effect and random-effect estimators, our research interest in the posting trends of time-invariant demographic groups restricts us to the use of the random-effect estimator, which does not rely solely on within-group variation. For this reason, the models are also all controlled for differences in the demographic variables to avoid spurious correlations caused by, for example, overrepresentation of city residents in the higher educational groups solely driving a false trend between education and profile posts. Further, deviating from the assumptions of the Poisson distribution that the variance equals the mean, cluster-robust standard errors are used to account for the significant amount of overdispersion found in our sample data – as accounted for in the section on posts (*ibid.*). Finally, the estimated regression coefficients are exponential to ease the interpretation as incidence rate ratios with multiplicative effects.

To test the percentage of posters in relation to non-posters over time according to age, the study reuses the dynamically adjusted age data set and present a descriptive overview of the percentage of posters in our data set over time.

Results

Change in posting frequency over time in relation to demographics

To answer the question of whether there is a significant change in posting frequency over time and whether it varies along with demographic variables (age, gender, education and area of residence), a Poisson regression analysis of the profile posts over time is conducted after normalization (e.g. deleting automated posts from non-whitelisted apps). The results from the regression analysis are shown in Table 2 (see next page).

Table 2 shows a significant increase in posts of 0.6 per cent per month over time for the total period from 2007 to 2014. This is expected from the existing literature (H1), as time spent on the internet appears to have been a predictor of participation in earlier studies along with a registered increase in reported posting behaviour across Pew surveys (Hargittai, 2007; Schradie, 2011). The total increase in monthly profile posts in the sample rose from 4 to 7 posts a month from 2007 to 2014. The development in the estimated population of Danes on Facebook is measured as being no higher than 0.8 per cent and no lower than 0.4 per cent (95% certainty).

Controlling for other demographic variables, Figure 2 shows that there is a high correlation between the y-intercept and the increase level in the data set:

This means that, despite finding the expected inequality of women posting more than men, inequality diminishes in our data set over time for gender (H3). However, the data set apparently shows inequality in age over time (H4). We see no inequality in medium and long education (short education is underrepresented), as expected (H2). Also as expected, the area of residence shows very little inequality and this inequality diminishes over time following the same slope as education (H5).

Returning to Table 2, this table shows significant trends for the gender, age and area of residence variables. Men have a larger increase in monthly profile posts than women, with a significant increase of one per cent compared with 0.2 per cent for women. When revisiting the presumption from the existing theory (McAndrew & Jeong, 2012) that women post more than men, this is true in our sample, especially at the beginning of the time period (men have a lower y-intercept). The study cannot infer to the general population that women post more at the end of the period due to the small difference; however, Figure 2 shows that the increase in the number of posts is greater for men than for women in the sample during the period.

Education as a demographic parameter shows no significant values, as expected, because the sample size is too small for short education. Therefore, the study cannot say anything significant about the influence of social economic status on posting inequality, despite the existing studies having pointed to this variable as a predictor of inequality. Furthermore, as the graph shows, there is almost no difference in the intercept and increase level in our data set for higher education and medium-length education.

The study shows a significant increase of 0.7 per cent for participants from rural areas, whereas there are no significant developments for participants from urban areas. The difference between rural and urban is smaller in the data set than that for gender.

Table 2. *Poisson regression of monthly wallposts over time (2007-2014) by gender, age, education and area of residence. Robust standard errors and exponentiated coefficients (incidence rate ratios) are used*

	Model 1 Coef./Std err.	Model 2 Coef./Std err.	Model 3 Coef./Std err.	Model 4 Coef./Std err.	Model 5 Coef./Std err.
Date (monthly)	1.006*** (0.001)	1.010*** (0.002)	1.003* (0.001)	1.000 (0.005)	1.007*** (0.001)
Gender (1=female)	1.471*** (0.159)	92.883*** (121.536)	1.447*** (0.153)	1.463*** (0.157)	1.445*** (0.156)
15-29	1.000 (.)	1.000 (.)	1.000 (.)	1.000 (.)	1.000 (.)
30-44	0.957 (0.126)	0.959 (0.119)	0.105 (0.161)	0.947 (0.129)	0.960 (0.123)
45-59	1.052 (0.195)	1.023 (0.177)	0.095 (0.199)	1.028 (0.196)	1.020 (0.182)
60+	1.239 (0.388)	1.100 (0.334)	0.001*** (0.002)	1.231 (0.381)	1.155 (0.350)
Short	1.000 (.)	1.000 (.)	1.000 (.)	1.000 (.)	1.000 (.)
Medium	1.469* (0.229)	1.496* (0.232)	1.459* (0.229)	0.058 (0.201)	1.481* (0.232)
Long	1.547** (0.247)	1.563** (0.246)	1.513** (0.243)	0.026 (0.087)	1.553** (0.248)
Urban (1=urban)	1.166 (0.116)	1.152 (0.113)	1.165 (0.115)	1.160 (0.116)	10.214 (13.109)
Gender:					
Male x Date		1.000 (.)			
Female x Date		0.993** (0.002)			
Age:					
15-29 x Date			1.000 (.)		
30-44 x Date			1.004 (0.002)		
45-59 x Date			1.004 (0.003)		
60+ x Date			1.011*** (0.003)		
Education:					
Short x Date				1.000 (.)	
Medium x Date				1.005 (0.006)	
Long x Date				1.007 (0.005)	
Urbanization:					
Rural x Date					1.000 (.)
Urban x Date					0.997 (0.002)
Constant	0.078*** (0.053)	0.006*** (0.006)	0.423 (0.403)	3.070 (10.050)	0.030*** (0.027)
N. of obs.	48,300	48,300	48,300	48,300	48,300

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Comment: The models are based on random-effects Poisson regression models to accentuate trends within time-invariant predictor variables (e.g. gender). Hausman tests comparing the consistency of random-effects to fixed effects estimates indicate discrepancies in random-effects estimates. Robust standard errors are used to account for overdispersion. $N = 922$.

Figure 2. Monthly posts per gender (top left), age (top right), education (bottom left) and area of residence (bottom right)

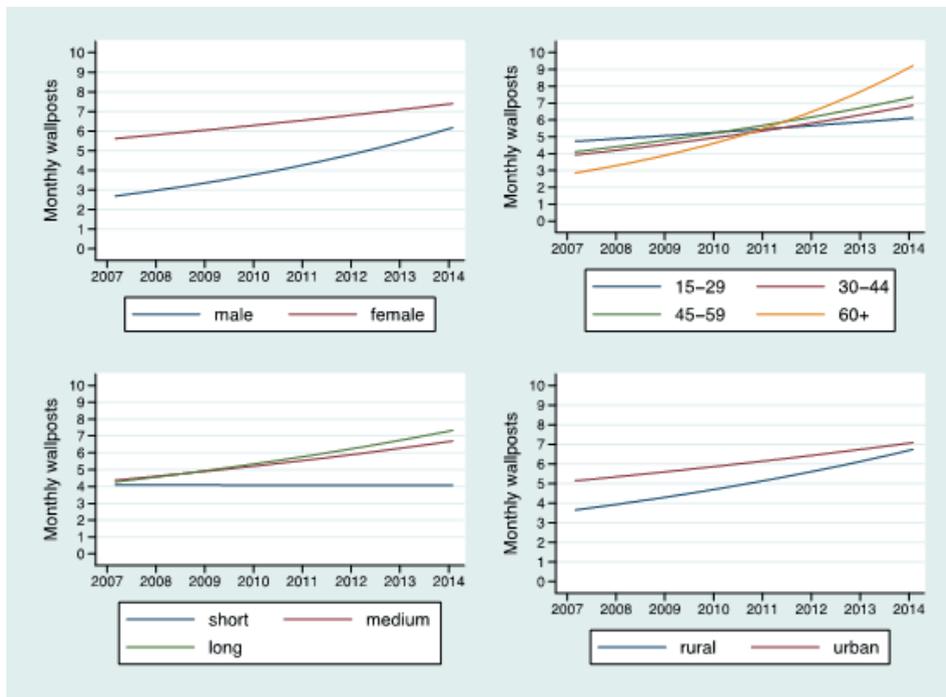


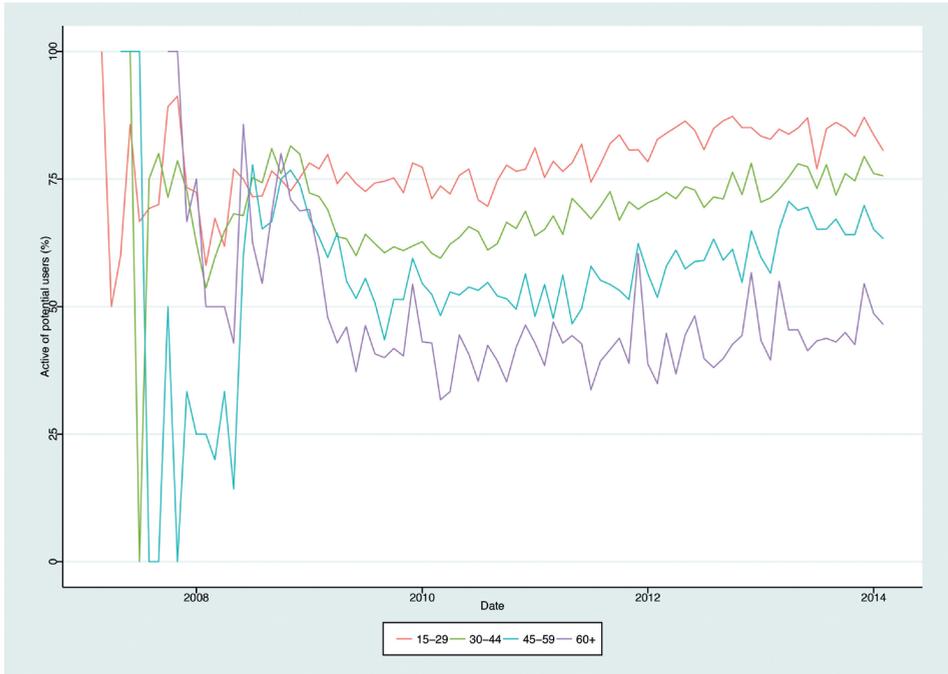
Figure 2 shows a larger average y-intercept for urban people, while the increase level for rural people is higher.

According to inequality in age as a demographic variable, all the age groups have an increase in monthly profile posts over time, as shown in Figure 2, increasing steadily from younger to older people. However, in Table 2, only young (15-29) and older (60+) participants show significant increases of 0.3 per cent and 1.4 per cent, respectively. Figure 2 also shows that, in the data set, there is a larger increase rate in the number of posts per month for older people than for younger people on the Facebook profile. When we look at age in isolation in our data set, the older participants have the lowest y-intercept at the beginning of the period and the youngest group has the highest. However, the older participants have the highest increase level and the youngest groups have the lowest increase level: both are statistically significant, despite the different entry points, and, due to the difference in the degree of increase in the period, the lines cross in 2011/12. However, despite these patterns in our data set, there is no significant pattern showing that young people post less than the older age groups in the total Danish Facebook population. If we isolate the last year to try to explain this trend (see the analysis in the appendix), the regression shows a 2.3 per cent increase in monthly profile posts among the 60+ category and an insignificant decrease tendency among the young category (15-29), although the monthly posting frequency for young people within the sample decreases by 0.7 per cent (compared with the increase level of 0.3% for the total time period).³

Posters and non-posters over time in relation to age

Turning to the next potential inequality between non-posters and posters over time (RQ2), we see large differences depending on the age group.

Figure 3. *The percentage of posters in age groups*



In all the age groups, the percentage of posting users (profile posts) is much higher from mid-2008 onwards. The large fluctuations in the graph before 2008 are due to a small number of participants. Before October 2007, there are only 100 participants in the data set. If we look at each age category and find a shared starting point at which there are at least 100 participants in each group, the intercept is January 2009. At this point, the percentage of users who make at least one monthly post is higher than 20 per cent for all the age groups. As Figure 3 shows, the age groups have almost the same ratio during the whole period. The 60+ group has the lowest percentage, between 25 per cent and 60 per cent, followed by the age group 45-59 with 45-75 per cent, then the age group 30-44 with 70-85 per cent, and finally the age group 15-29 with the highest percentage of posters, 70-90 per cent. At any given time, we find that posters make up *more* than 40 per cent on average of the total number of potential users, but, as expected within this data set and with this method, we can conclude that the representation of different age groups varies considerably in the digital trace data, with the percentage of younger people being notably higher than that of older people (H6).

The analysis presented here only accounts for content uploaded within the profile feed, excluding auto-generated posts. Furthermore, the remaining users may post updates in groups or use Messenger as a way to be active on Facebook. Such usage is not accounted for in this article.

Discussion

What are the implications of the results in this article compared with the previous studies? The findings from the API trace data in this article in many ways supplement the existing knowledge on posting behaviour from the survey studies that coined the basis for the hypotheses. The article finds an overall increase in posting frequency on Facebook, despite a trend in the data set of younger people posting less towards the end of the period. It is very important for the research field not to overinterpret the possible decrease in profile posting frequency among the young age group. The tendency shown in this study is not significant and therefore cannot confidently support this tendency.

At the same time, the study documents inequality, especially in age and gender, although these diminish over time for gender in the data set. The study finds inequality in age in the ratio between posters and non-posters. Overall, the sensitivity to the time aspects adds to the existing knowledge in the field of social media behaviour, as it shows that, if we only take into consideration a single-standing slice of time data set, then we will not discover these trends.

Furthermore, the inequality that the study shows in posting behaviour has two consequences for the initial motivation of this article. First, it points to inequality in the *potential* for different demographics to set the frame (Scheufele, 1999), for example for gender and age, although the analytical method of this article did not allow details of the actual frame-setting patterns to be identified; this is left for future studies to pursue through Natural Language Processing (NLP) methods. Second, the findings on inequality both in frequency and in poster ratio support the existing scholarly critique of treating big data as a “god’s eye view” (Blank, 2017; Blank & Lutz, 2017; Hargittai, 2015; Kitchin, 2014), as they show that not all sociodemographic profiles are equally visible in the digital trace data through posts both on demographics and within demographics over time. This means that, when we use such digital trace data to make inferences for a larger population, we need not only to consider sampling biases (Lomborg & Bechmann, 2014) but also to take both time and demographic categories into consideration to balance our samples.

The data set in particular shows unequal posting behaviour for age and gender, which is supported by the existing survey studies but again highlights the need to investigate different demographics and broader samples than, for instance, just college students (Hargittai & Walejko, 2008). McAndrew and Jeong (2012) suggested that women post more than men, but the study presented in this article shows a larger increase level in monthly posts among men. Ertiö, Kukkonen and Räsänen (2018) and Young (2011) showed that around 40 per cent of users make posts, but the average poster percentage is higher in this data set and, when broken down into different age groups, the study shows very large differences in the poster ratio (defined here as posters/non-posters); the younger the individual, the higher the ratio. Whereas Nonnecke and Preece (2000) described large inequality in the lurker/poster ratio in different communities, the results of this study contribute to the field by stabilizing the community but showing inequality in the demographic variable age. Even though the ratio of older posters is lower, the posting frequency increase level is higher for the older people’s category (Figure 2), which can indicate that a smaller percentage of older posters drives the frequency up by posting more.

Limitations

Sampling biases have the potential to influence the results as our sample may contain frequent posters, because our strata do not take this into account and the use of an internet panel may encourage the selection of such users (Baker et al., 2010). However, there is no immediate reason to believe that this sample bias varies across age groups.

Another limitation of the study is that the data set does not account for potential decreases and subsequent changes in participation inequality that may have happened after 2014. However, by studying the developments over seven years with a focus on the general research interest in participation inequality, the findings are still relevant as a much-needed longitudinal contribution to participation inequality research (Hargittai & Walejko, 2008).

The results are not immediately generalizable to other nationalities due to varying media systems (Hallin & Mancini, 2004). This is both a strength and a weakness. It provides us with a new national case study to supplement the existing studies, which are primarily from the US, Australia and the UK. At the same time, it calls for further research to apply similar approaches to the same or related research questions to maintain a focus on robust demographic predictors within each country yet compare them across countries. This is especially interesting for countries with a similar Facebook penetration rate across age groups but with different gender roles, as for instance accounted for in the global gender gap index (Hausmann et al., 2012).

Similarly, future research on posting behaviour on Facebook through digital trace data may cover other relevant predictors that are not accounted for in this study, for instance online skills (Blank, 2013) and the duration of community membership.

Conclusion

This article found a significant increase in the amount of monthly posts on Facebook from 2007 to 2014. Furthermore, there was no significant decrease in posts for young people or any other age group or isolated demographic parameter (gender, education or area of residence) in the Poisson regression, although the sample showed a decreasing trend in the last year. Women posted more at the beginning of the period in our sample, but the significant increase level was larger for men than for women, which means that the inequality in our sample diminished over time. Furthermore, the study showed notable inequality in posting frequency for age over time and for the ratio between monthly posting users and non-posting users, with the highest percentage of posting users being in the youngest age group.

Despite finding a higher percentage of posters than the previous survey studies, the study found inequality in the demographic profile of those who are mostly represented in the post data and thereby both potentially violated privacy but also set the frame. The findings largely confirm the trends in the prior survey studies on participation inequality on Facebook but provide nuances to these studies both by providing a new data source through log/trace data and by incorporating time as a variable that shows how the inequality develops. The time aspect in the study on inequality both in posting frequency and in poster ratio is important to notice when raw data (Bowker, 2014) are treated as an objective indicator of “true” knowledge and meaning for a community or

society. If we do not take time into account, we become blind to how changes happen in the participation inequality.

Funding

The work was generously funded by the Aarhus University Research Foundation [grant number AUFF-E-2015-FLS-8-55], and the article was finalized at the Aarhus Institute of Advanced Studies with a Jens Christian Skou Research Fellowship.

Notes

1. Our API data collection differed from the Gemius survey methods; for example, we also measured mobile-only users but Gemius did not. This is one of the reasons for our choice not to map the Gemius population in our inferences but to map according to the sample strata division estimates and discuss the implications of this.
2. The area of residence is divided into urban and rural, in which urban is defined as living in or around (suburban areas) the five largest cities (>70,000 citizens) of Denmark. Race has frequently been included in American surveys on digital inequality, but ethnic minorities have a much more diversified profile in Denmark and account for under 10 per cent of the total population. Therefore, while we included ethnic minorities in our sample, we did not use them as a demographic variable in the sampling or the analysis (Statistics Denmark).
3. In the more optimistic negative binominal regression, this decrease is significant.

References

- Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., ... Zahs, D. (2010). AAPOR report on online panels. *Public Opinion Quarterly*, 74(4): 711-781.
- Baumer, E. P. S., Adams, P., Khovanskaya, V. D., Liao, T. C., Smith, M. E., Schwanda Sosik, V. & Williams, K. (2013). Limiting, leaving, and (re)lapsing: An exploration of Facebook non-use practices and experiences. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 3257-3266). New York, NY: ACM. <https://doi.org/10.1145/2470654.2466446>
- Bechmann, A. (2014). Non-informed consent cultures: Privacy policies and app contracts on Facebook. *Journal of Media Business Studies*, 11(1): 21-38. <https://doi.org/10.1080/16522354.2014.11073574>
- Bechmann, A. (2015). Managing the interoperable self. In *The ubiquitous Internet: User and industry perspectives*. New York: Routledge. Retrieved from <https://www.taylorfrancis.com/books/e/9781317931409/chapters/10.4324%2F9781315856667-10>
- Bechmann, A. (2019). Data as humans: Representation, accountability, and equality in big data and machine learning. In *Power and rights in the online domain*. Cambridge, MA: MIT Press.
- Bechmann, A. & Lomborg, S. (2013). Mapping actor roles in social media: Different perspectives on value creation in theories of user participation. *New Media & Society*, 15(5): 765-781. <https://doi.org/10.1177/1461444812462853>
- Bechmann, A. & Vahlstrup, P. B. (2015). Studying Facebook and Instagram data: The Digital Footprints software. *First Monday*, 20(12): 1-13. <https://doi.org/10.5210/fm.v20i12.5968>
- Bernstein, M. S., Bakshy, E., Burke, M. & Karrer, B. (2013). Quantifying the invisible audience in social networks. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 21-30). New York, NY: ACM. <https://doi.org/10.1145/2470654.2470658>
- Blank, G. (2013). Who creates content? *Information, Communication & Society*, 16(4): 590-612. <https://doi.org/10.1080/1369118X.2013.777758>
- Blank, G. (2017). The digital divide among Twitter users and its implications for social research. *Social Science Computer Review*, 35(6): 679-697. <https://doi.org/10.1177/0894439316671698>

Acknowledgements

The author would like to thank Userneeds, the participants who shared their data for the purpose of research, student assistant Anders Geil, who carried out the regression analyses, Peter Vahlstrup, who co-developed the Digital Footprints software, Stine Lomborg, Katrin Tiidenberg and anonymous reviewers for their insightful suggestions.

- Blank, G. & Lutz, C. (2017). Representativeness of social media in Great Britain: Investigating Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram. *American Behavioral Scientist*, 61(7): 741-756. <https://doi.org/10.1177/0002764217717559>
- Bowker, G. C. (2014). Big data, big questions: The theory/data thing. *International Journal of Communication*, 8(0): 5.
- Bruns, A., Bechmann, A., Burgess, J., Chadwick, A., Clark, L. S., Dutton, W. H., ... Zimmer, M. (2018). Facebook shuts the gate after the horse has bolted, and hurts real research in the process. *Internet Policy Review*. Retrieved from <https://policyreview.info/articles/news/facebook-shuts-gate-after-horse-has-bolted-and-hurts-real-research-process/786>
- Burke, M., Marlow, C. & Lento, T. (2010). Social network activity and social well-being. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1909-1912). New York, NY: ACM. <https://doi.org/10.1145/1753326.1753613>
- Cameron, A. C. & Trivedi, P. K. (2010). *Microeconometrics using Stata: Revised edition* (2nd edition). College Station, TX: Stata Press.
- Correa, T. (2010). The participation divide among "online experts": Experience, skills and psychological factors as predictors of college students' web content creation. *Journal of Computer-Mediated Communication*, 16(1): 71-92. <https://doi.org/10.1111/j.1083-6101.2010.01532.x>
- Deursen, A. J. A. M. van & Dijk, J. A. G. M. van. (2015). Internet skill levels increase, but gaps widen: A longitudinal cross-sectional analysis (2010-2013) among the Dutch population. *Information, Communication & Society*, 18(7): 782-797. <https://doi.org/10.1080/1369118X.2014.994544>
- Ertiö, T., Kukkonen, I. & Räsänen, P. (2018). Social media activities in Finland: A population-level comparison. *Convergence*, <https://doi.org/10.1177/1354856518780463>
- Hallin, D. C. & Mancini, P. (2004). *Comparing media systems: Three models of media and politics*. Cambridge, UK: Cambridge University Press.
- Hargittai, E. (2007). Whose space? Differences among users and non-users of social network sites. *Journal of Computer-Mediated Communication*, 13(1): 276-297. <https://doi.org/10.1111/j.1083-6101.2007.00396.x>
- Hargittai, E. (2015). Is bigger always better? Potential biases of big data derived from social network sites. *ANNALS of the American Academy of Political and Social Science*, 659(1): 63-76. <https://doi.org/10.1177/0002716215570866>
- Hargittai, E. & Walejko, G. (2008). The participation divide: Content creation and sharing in the digital age. *Information Communication and Society*, 11(2): 239-256. <https://doi.org/10.1080/13691180801946150>
- Hausmann, R., Tyson, L. D. & Zahidi, S. (2012). *The global gender gap report 2012*. Geneva: World Economic Forum.
- Hoffmann, C. P., Lutz, C. & Meckel, M. (2015). Content creation on the Internet: A social cognitive perspective on the participation divide. *Information, Communication & Society*, 18(6): 696-716. <https://doi.org/10.1080/1369118X.2014.991343>
- Junco, R. (2013). Comparing actual and self-reported measures of Facebook use. *Computers in Human Behavior*, 29(3): 626-631. <https://doi.org/10.1016/j.chb.2012.11.007>
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. London: Sage.
- Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A. & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, 30(4): 330-342. <https://doi.org/10.1016/j.socnet.2008.07.002>
- Lomborg, S. & Bechmann, A. (2014). Using APIs for data collection on social media. *The Information Society*, 30(4): 256-265. <https://doi.org/10.1080/01972243.2014.915276>
- Lutz, C. (2016). A social milieu approach to the online participation divides in Germany. *Social Media + Society*, 2(1) :<https://doi.org/10.1177/2056305115626749>
- Lutz, C., Hoffmann, C. P. & Meckel, M. (2014). Beyond just politics: A systematic literature review of online participation. *First Monday*, 19(7). <https://doi.org/10.5210/fm.v19i7.5260>
- Marwick, A. E. & Boyd, Danah. (2014). Networked privacy: How teenagers negotiate context in social media. *New Media & Society*, 16(7): 1051-1067. <https://doi.org/10.1177/1461444814543995>
- McAndrew, F. T. & Jeong, H. S. (2012). Who does what on Facebook? Age, sex, and relationship status as predictors of Facebook use. *Computers in Human Behavior*, 28(6): 2359-2365. <https://doi.org/10.1016/j.chb.2012.07.007>
- Nonnecke, B. & Preece, J. (2000). Lurker demographics: Counting the silent. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 73-80). New York, NY: ACM. <https://doi.org/10.1145/332040.332409>
- Scheufele, D. A. (1999). Framing as a theory of media effects. *Journal of Communication*, 49(1): 103-122. <https://doi.org/10.1111/j.1460-2466.1999.tb02784.x>

- Schradie, J. (2011). The digital production gap: The digital divide and Web 2.0 collide. *Poetics*, 39(2): 145-168. <https://doi.org/10.1016/j.poetic.2011.02.003>
- Schradie, J. (2013). Big data not big enough? How the digital divide leaves people out. Retrieved from <http://mediashift.org/2013/07/big-data-not-big-enough-how-digital-divide-leaves-people-out/> [Accessed 2018, November 8].
- Van Dijk, J. A. G. M. (2006). Digital divide research, achievements and shortcomings. *Poetics*, 34(4): 221-235. <https://doi.org/10.1016/j.poetic.2006.05.004>
- Wijas-Jensen, J. (2014). *It-anvendelse i befolkningen 2014*. Copenhagen: Statistics Denmark. Retrieved from <https://www.dst.dk/da/Statistik/Publikationer/VisPub?cid=18686>
- Young, K. (2011). Social ties, social networks and the Facebook experience. *Australian Journal of Emerging Technologies and Society*, 9: 20-34.

Appendix A

Poisson regression of monthly wallposts over time (2013-2014) by gender, age, education and urbanization. Robust standard errors and exponentiated coefficients (incidence rate ratios) are used

	Model 1 Coef./Std err.	Model 2 Coef./Std err.	Model 3 Coef./Std err.	Model 4 Coef./Std err.	Model 5 Coef./Std err.
Date (monthly)	1.000 (0.003)	0.996 (0.006)	0.993 (0.005)	1.002 (0.013)	1.002 (0.004)
Gender (1=female)	1.103 (0.120)	0.017 (0.072)	1.085 (0.119)	1.105 (0.120)	1.105 (0.121)
15-29	1.000 (.)	1.000 (.)	1.000 (.)	1.000 (.)	1.000 (.)
30-44	0.840* (0.069)	0.847* (0.070)	0.481 (2.411)	0.842* (0.069)	0.839* (0.069)
45-59	0.846 (0.107)	0.850 (0.106)	0.000 (0.002)	0.085 (0.106)	0.845 (0.106)
60+	0.620*** (0.086)	0.623*** (0.086)	0.000* (0.000)	0.625*** (0.088)	0.624*** (0.087)
Short	1.000 (.)	1.000 (.)	1.000 (.)	1.000 (.)	1.000 (.)
Medium	1.666** (0.290)	1.663** (0.290)	1.670** (0.294)	0.344 (3.241)	1.665** (0.290)
Long	1.817*** (0.309)	1.813*** (0.307)	1.795*** (0.307)	45.064 (407.550)	1.819*** (0.309)
Urban (1=urban)	0.982 (0.110)	0.983 (0.110)	0.987 (0.113)	0.982 (0.110)	32.477 (147.544)
Gender:					
Male x Date		1.000 (.)			
Female x Date		1.007 (0.007)			
Age:					
15-29 x Date			1.000 (.)		
30-44 x Date			1.001 (0.008)		
45-59 x Date			1.012 (0.008)		
60+ x Date			1.031* (0.013)		
Education:					
Short x Date				1.000 (.)	
Medium x Date				1.002 (0.015)	
Long x Date				0.995 (0.014)	
Urbanization:					
Rural x Date					1.000 (.)
Urban x Date					0.995 (0.007)
Constant	5.050 (10.650)	57.825 (211.587)	489.468 (1689.259)	1.328 (11.516)	1.382 (3.296)
N. of obs.	12,276	12,276	12,276	12,276	12,276

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Comment: The models are based on random-effects Poisson regression models to accentuate trends within time-invariant predictor variables (e.g. gender). Hausman tests comparing the consistency of random effects to fixed effects estimates indicate discrepancies in random effects estimates. Robust standard errors are used to account for overdispersion. $N = 922$.