Close Social Networks Among Older Adults

Gil-Clavel, Sofia; Zagheni, Emilio; Bordone, Valeria

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

ORIGINAL RESEARCH

Close Social Networks Among Older Adults: The Online and Offline Perspectives

Sofia Gil-Clavel[1,2] · **Emilio Zagheni[1]** · **Valeria Bordone[3]**

Abstract

Qualitative studies have found that the use of Information and Communication Technologies is related to an enhanced quality of life for older adults, as these technologies might act as a medium to access social capital regardless of geographical distance. In order to quantitatively study the association between older people's characteristics and the likelihood of having a network of close friends offline and online, we use data from the Survey of Health, Ageing and Retirement in Europe and data from Facebook. Using a novel approach to analyze aggregated and anonymous Facebook data within a regression framework, we show that the associations between having close friends and age, sex, and being a parent are the same offline and online. Migrants who use internet are less likely to have close friends offline, but migrants who are Facebook users are more likely to have close friends online, suggesting that digital relationships may compensate for the potential lack of offline close friendships among older migrants.

Keywords Social networks · Social capital · Older people · Social media data · Europe

Introduction

Over the last decade, there has been a rapid increase in the possibilities offered by online environments to cultivate meaningful social relationships. As more and more areas of life are "digitized," this process—often referred to as "digitalization of life"—creates both challenges and opportunities. On the one hand,

✉ Sofia Gil-Clavel
  gil@demogr.mpg.de

[1] Laboratory of Digital and Computational Demography, Max Planck Institute for Demographic Research, Rostock, Germany

[2] Faculty of Spatial Sciences, University of Groningen, Groningen, The Netherlands

[3] Department of Sociology, University of Vienna, Vienna, Austria

unequal access to digital technologies and heterogeneous levels of digital literacy may amplify existing inequalities. On the other hand, for some socio-demographic groups, access to digital resources may help compensate for lower levels of social capital and serve as an equalizer, thus, reducing overall inequalities in areas such as social support.

Older adults are part of a key demographic group that could potentially benefit largely from access to digital technologies, but that is also at risk of being excluded from reaping the gains of a digital world if they fall on the "wrong" side of the digital divide. Understanding the role of internet and Social Network Sites (SNSs) in later life requires a broad and comparative perspective. In 2018, in Europe, 75% of adults aged 55–64 declared to have used internet in the last 3 months, whereas for those over 65, the percentage dropped to 56% (Eurostat, 2018). The use of Information and Communication Technologies by older adults—internet among them—is related to an enhanced quality of life (Francis et al., 2019; Sims et al., 2016). This relationship is believed to result from the access to social capital that these technologies offer (Neves, 2013). In particular, SNSs, such as Facebook and Twitter, play an important role among the Information and Communication Technologies resources that older people have access to, because they help older people to overcome perceptions of social isolation and loneliness (Jung & Sundar, 2016; Ballantyne et al., 2010).

Demographers have used survey data to study population dynamics since the advent of the discipline. But there are certain populations that are still difficult to sample. These include, among others, migrants (Beauchemin & González-Ferrier, 2011) and older migrants in particular (Warnes & Williams, 2006). The digital revolution has created new opportunities to passively collect socio-demographic and behavioral data through social media platforms like Facebook and Twitter (Alburez-Gutierrez et al., 2019; Edelmann et al., 2020; Lazer & Radford, 2017). Even though these platforms were not conceived for research purposes, the fast growth of their worldwide user base has led researchers to consider them as a complementary data source for demographic research.

Facebook use can be considered a prime example of SNS, given that it is the most frequently used SNS worldwide with around 1.62 billion daily active users (Facebook Inc., 2019). The use of these data has advantages for demographic research. For example, it offers ways to obtain information about demographic characteristics and interests of subpopulations that otherwise would be difficult to reach and to study. Thus, Facebook data have already been used to study access to digital technologies (Fatehkia et al., 2018), immigrant's cultural assimilation (Dubois et al., 2018; Stewart et al., 2019), and to estimate migrant stocks across countries (Zagheni et al., 2017).

To the best of our knowledge, no study has investigated the use of Information and Communication Technologies in older populations using a combination of social media and more traditional survey data. Here, we use both traditional and new sources of data: The Survey of Health, Ageing and Retirement in Europe (SHARE) and Facebook data, respectively. These are used to expand the literature regarding (1) the representativeness of Facebook data for aging research and (2) the association between older people's characteristics and having a close network of friends

offline and online. This way we aim to offer new insights into who has access to the benefits that social capital could impart through the use of internet and SNS.

This paper aims to contribute to the literature both methodologically and substantively. First, we show how data from the Facebook Marketing Application Programming Interface (API) can be used to understand relationships in a way that mimics micro-level regression analysis, even though these data come only in aggregate form. Second, we demonstrate how digital trace data can provide new insights into the use of SNS by older people.

In what follows, we first present the theoretical background of our work. We discuss the definition of social capital and its determinants; and the relation between social capital, internet and SNS use, and older adults' health. Then we give a brief explanation of the variables for which we expect to find some heterogeneity among the offline and online groups in relation with having close friends. Next, we introduce the two databases we use for our analyses and explain the methodological approach. After showing the results obtained from the analyses of each database, we summarize the findings and conclude.

Background

The concept of social capital as used in current literature derives from different theoretical traditions (see e.g., Putnam, 1993; Coleman, 1988; Bourdieu, 1986) that have in turn determined a variety of methodologies with which the concept has been studied across social sciences. In this work, we rely on the definition of social capital proposed by Lin (1999, p. 39): "investment in social relations by individuals through which they gain access to embedded resources to enhance expected returns of instrumental or expressive actions." Here, expressive actions are resources already possessed by the person, such as physical health, mental health, and life satisfaction. This return is more likely to be mobilized in denser networks with more intimate and reciprocal relations among members (Lin, 1999). Therefore, we use close friends as a proxy to social capital.

Following the fast and global spread of digital technology, research has started to explore whether and how the use of Information and Communication Technologies, in particular internet and SNS, could help older adults to improve their health. The internet may act as a medium for older adults to achieve better health through access to information and social relationships (Rios et al., 2019; Sum et al., 2008). It may also help to maintain close relationships (i.e., bonding social capital) (Neves and Barbara, 2015). Similarly, the use of SNS represents an accessible and relatively low-cost mechanism to enhance social connections at older ages (Vitak, 2014). For example, it has been shown that SNS can reduce loneliness experienced at particular moments of the day and related to not being part of a community (Ballantyne et al., 2010). SNS also allow older people to communicate with family and acquaintances and to receive support from them (Lee et al., 2013).

Our study explores the heterogeneity among different groups of people that have close friends in relation to whether they do or do not use Information and Communication Technologies. More specifically, it focuses on variables that, according to

the literature, are associated with the adoption of Information and Communication Technologies among older people: sex, age group, level of education, being a parent, and being a migrant. In the case of sex, research shows that older women benefit the most from the use of Communication Technologies, as they have higher contact and exchange of emotional support with their adult children as compared to men (Peng et al., 2018; Suitor et al., 2016). Comparison of older adults by age is important because people aged 65 + tend to use Information and Communication Technologies less compared also to the immediately below 50–64 age group (Hunsaker & Hargittai, 2018). This comparison is also important as it marks important transitions in the individuals' lives ("grey divide"), such as retirement, which are often associated with a decrease in the use of Information and Communication Technologies (Friemel, 2016). Research also highlights that the use of Information and Communication Technologies by older adults highly depends on their level of education, with people with higher levels of education using the internet more compared to those with lower levels of education (Hargittai, 2020; Hunsaker & Hargittai, 2018; Lee et al., 2011). We additionally focus on whether older adults have children because Information and Communication Technologies are used to connect with geographically distant kin (Quan-Haase et al., 2018). Qualitative studies also suggest that migrants tend to rely on Information and Communication Technologies to exchange emotional support over distance (Baldassar et al., 2016; Bates & Komito, 2012; Komito, 2011) and that older adults find Communication Technologies useful when relatives live abroad or far away (Neves et al., 2018).

In this article, we build on the literature on social capital and population aging. Following Lin (1999), we consider having close friends as a proxy for access to social capital and its expected returns, such as mental health and life satisfaction. We assess, at the population level, which groups of older people are more likely to have close friends depending on whether they are internet users, Facebook users (as a proxy for SNS usage), or offline. Implicitly, we rely on Lin's (1999) prediction that cyber networks create social capital, because they offer free access to information, data, and other individuals, in a way that transcends time and space.

Data

We use two sources of data to analyze and contrast the characteristics of people's online and offline social networks. Specifically, we draw on data from SHARE and from the Facebook Marketing API.

SHARE Database

SHARE is a longitudinal survey representative of the population aged 50 years and over in Europe. The SHARE survey design enables scientists to draw inferences about the population of 50 years and older across countries by using probability-based sampling. For this study, we use wave 6, conducted in 2015 in 17 European countries (Appendix B) with 66,153 participants, because it is the most recent wave

including the module on Social Networks. The target population of SHARE wave 6 "[…] consists of persons born in 1964 or earlier, and persons who are a spouse/partner of a person born in 1964 or earlier, who speak (one of) the official language(s) of the country (regardless of nationality and citizenship) and who do not live either abroad or in institutions such as prisons and hospitals during the entire fieldwork period" (Bergmann, De Luca, and Scherpenzeel 2017, 77).

For a complete description of the variables used in the study, the reader can refer to Appendix A of this paper. Here, we briefly summarize them. (1) *Sex* is a dummy variable with 1 meaning the person is a woman and 0 a man. (2) *Age* is a categorical variable that can be either 0 if the respondent's age is 50–64 or 1 if 65+. (3) *Education* is a categorical variable that can take values 0 if the respondent's highest educational attainment is below college degree, 1 if it is college or above, and 2 if unspecified. (4) *Parent* is a dummy variable with 1 meaning that the person has at least one child and 0 otherwise. (5) *Immigrant* is a dummy variable taking value 1 if the respondent was not born in the country of interview and 0 otherwise. (6) *Friend* is a dummy variable taking value 1 when the respondent declares to have at least one person with whom they feel somewhat close, very close, or extremely close and 0 otherwise. (7) *Internet* is a dummy variable indicating whether the respondent used the internet at least once during the previous 7 days, for e-mailing, searching for information, making purchases, or for any other purpose.

Given the relevance of the variable *Friend* for our study, we now provide further details about it. The Social Network Module of SHARE is based on "[…] a name generating mechanism in which respondents identify the people who are important to them and then subsequently add information on each person (up to seven named)" (Schwartz, Litwin, and Kotte 2017, 22). This information includes the ties that were involved in social exchange (e.g., the financial or time transfers in which people engage). The name generating mechanism is a common strategy to learn about people's ego networks. This approach is suggested when the goal is to estimate the benefits of being part of a network (Merluzzi & Burt, 2013), or to compare outcomes across surveys, as it helps standardize results (Maya Jariego 2018). However, researchers have also highlighted that this practice can bias the results, when estimating the total number of links (friends, colleagues, etc.) that a person has (Goodreau et al., 2009; Neal & Neal, 2017). Based on SHARE data, older people mentioned on average 2.3 names, and only 5% of the sample mentioned six or more names. Therefore, we believe that, for the networks considered in this work, the name generating mechanism should not bias the results, as older people's networks tend to be small.

As described by Schwartz, Litwin, and Kotte (2017), the SHARE interview of this module starts with the question "Over the last 12 months, who are the people with whom you most often discussed important things?" to which the interviewee can answer with up to six names and name one additional person that is important for them "for some other reason." Afterwards, more details are asked about the named persons, if such information does not appear elsewhere in the interview (e.g., children's data). The information includes gender, year of birth, occupational status, and marital status of each mentioned person, as well as their residential proximity, frequency of contact, and emotional closeness to them. In particular, based on the

answer ((1) Not very close; (2) Somewhat close; (3) Very close; and (4) Extremely close) to the question "How close do you feel to [mentioned name]?" we built the dummy variable *Friend*, taking value 1 when the respondent declares to have at least one person with whom they feel somewhat close, very close, or extremely close and 0 otherwise. We decided to aggregate those categories to facilitate comparability with the data from Facebook which have only two categories. After performing a sensitivity analysis, the results from SHARE do not change if the category "Somewhat close" is included in the category taking value 1 or in the one taking value 0.

Facebook Database

The Facebook Marketing API is a tool that allows access to the Facebook Adverts Manager in a programmatic way. The Facebook Adverts Manager platform gives advertisers the approximate number of Facebook users that match certain characteristics, before an ad is launched and before any payment is performed or requested (for a detailed description the reader can refer to Zagheni et al. (2017)). This platform was built for marketing purposes, but demographers and sociologists, among others, have found an invaluable source of information in it.

We use the Facebook API version 3.2[1] to retrieve the "Daily Active Users" during one month—from June 9 to July 9 2019—that matched the combination of the following characteristics[2]: (0) Live in one of the 17 countries showed in Appendix B (the same as for SHARE); (1) *Sex*: Declared gender is either female or male; (2) *Age*: Declared age is either between ages 50–64 or 65 +; (3) *Education*: Declared education is Below College, College Or Above, or Unspecified; (4) *Parent*: Classified as parents, coded as 1 = yes and 0 = no; (5) *Immigrant*: Classified as immigrants, coded as 1 = yes and 0 = no; (6) *Friend*: Classified as being close friends of people with birthdays in a month, coded as 1 = yes and 0 = no. The detailed description of the variables can be found in Appendix A.

Facebook does not specify whether the variables 4–6 come from users' self-declared characteristics or whether Facebook classifies the users based on their networks or other data.[3] Here, we summarize some articles published by Facebook that shed some light on this. Articles based on the Facebook population, rather than surveys, show that Facebook researchers consider users' self-declared characteristics, but they also highlight the features of the users' networks. Backstrom et al. (2011) propose a measure for the analysis of personal networks, based on the way individuals divide their attention across contacts. Their metrics consider different modalities that can be summarized as communication and viewing based that are used to rank users' close friends. In the case of parent–children relations, Burke et al. (2013, p. 4) show that "Overall, 37.1% of English-speaking, monthly active US Facebook users have specified either a parent or child relationship on the site"; that children and

---

[1] https://developers.facebook.com/docs/graph-api/changelog/version3.2/. Accessed 30 September 2020.

[2] These subsets are mutually exclusive.

[3] https://www.facebook.com/ads/about/?entry_product=ad_preferences. Accessed 30 September 2020.

parents tend to befriend the same family members on Facebook, as well as some of the children's friends; and that their type of communication differs from the communication with non-nuclear family. Regarding immigrants, Herdağdelen et al. (2016) analyze users in the United States that specified home town (home country) in a country different from United States. In order to increase the reliability of the dataset, they constrain the sample to those with at least two friends currently living in their home country and another two friends currently living in the United States. Herdağdelen et al. (2016) also compare their results with US national statistics, showing that they are highly correlated.

Returning to our Facebook data, the total number of data points per country that we retrieved per day was $17 \times 2 \times 2 \times 3 \times 2^3 = 1632$. We did this for 31 days, resulting in a database with $31 \times (1632) = 50{,}592$ rows. Although Facebook returns both the daily and the monthly active users any time their data are retrieved, we use the Facebook Daily Active Users. This is because we are working with populations that can be smaller than 000 users and the Facebook Monthly Active Users value has a lower bound of 1000, whereas the Daily Active Users lower bound is 100.[4] In order to simplify notation, we will refer to the Facebook Daily Active Users as Facebook users.

Methodology

In order to study older people online and offline, first we need to assess whether Facebook users have demographic characteristics approximately similar to the ones of SHARE respondents who use internet. For this, we first compare both total and percentages of internet and Facebook users by demographic attributes. Assessing this is important to evaluate the extent of the bias when using Facebook users to approximate internet users in European countries.

Second, we check the proportions (P) of those older people in SHARE who declared to have used internet, against the proportions of older people in Facebook. We study the structure of the data by breaking it down into basic characterizations (Eq. 1). Specifically, we look at proportions of older people that are immigrant (*immigrant*), have close friends (*friend*), have children (*parent*), or none of the previous (*none*).

$$P_{ij}(s, a, e) = \frac{\#Users_{ij}(s, a, e)}{\sum_{s,a,e} \#Users_{ij}(s, a, e)} \tag{1}$$

In Eq. 1, *#Users* is the number of older people who use either internet (based on SHARE) or Facebook, represented by the index *i* and have one of the following characterizations: immigrant, friend, parent, or none, represented as the index *j*. We

---

[4] Al though the lower bound is 100, we also occasionally retrieve zeros when querying the API.

break down those groups by demographic characteristics: sex $s \in \{$Female, Male$\}$; age $a \in \{$50–64, 65+$\}$; and level of education $e \in \{$Unspecified, Below College, College or Above$\}$. This way, for example, the proportion of Facebook users that are mothers between 50 and 64, with an unspecified level of education would be given by

$$P_{Facebook,parent}(female, 50 - 64, Unspecified).$$

The analysis of these proportions helps us to study the association between the demographic distributions in these databases. In this case, we would expect to see a positive correlation between the internet and Facebook demographic proportions by characterization. A positive correlation means that the Facebook and internet populations are associated, and that an increase in the internet (Facebook) proportions is related to an increase in the Facebook (internet) ones. A negative correlation is not to be expected, and a correlation close to zero would mean that there is no association between these databases.

The second goal is to understand the association between older people's characteristics and their network of close friends, both for those who are online and those who are offline. For this, we use the same type of statistical analysis on two complementary datasets: non-internet users vs. internet users and non-internet users vs. Facebook users. We want to test whether the use of internet or Facebook has a differential effect on having close friends from those that are offline. For this, we test whether the coefficients of the model for non-internet users are statistically different from the coefficients from the internet and Facebook models.

According to Brame et al. (1998), we can test whether the coefficients of two identical generalized linear models are the same, when these ones are run in two independent groups. In this case, we assume that the non-internet group is independent from both the internet and Facebook groups. This test, which is performed with a z-score test (Brame et al., 1998), is also called Wald test for no difference in two independent samples. One of the novel aspects of our work is the use of classic statistical techniques to study aggregate counts, applied to Facebook users' data that anyone can obtain, at an aggregate level, from the Facebook Marketing API, but that are not available in the form of individual-level data as that would lead to privacy and ethical issues. More specifically, we test whether associations in the online world are statistically different from the offline ones.

The statistical analysis is based on logit models (Eq. 2) that are run independently in our databases: (1) SHARE non-internet users, (2) SHARE-Internet users, and (3) Facebook users.

$$logit(\pi_{friend_i}) = \alpha + \beta_1 age_i + \beta_2 sex_i + \beta_3 parent_i + \beta_4 immigrant_i. \quad (2)$$

In Eq. (2), i is an index that represents the data source for the model: (1) SHARE non-internet users, (2) SHARE-Internet users, and (3) Facebook users.

The dependent variable of the logit models is *friend*, a proxy for having close friends. For the Facebook data, we infer that close friend (*friend*) is a variable that is produced from the way Facebook users divide their attention across contacts (Backstrom et al., 2011), while for SHARE, it is whether the participant feels somewhat

close, very close, or extremely close to a person. The explanatory variables are *parent* and *immigrant*: in the case of Facebook data, we infer that these variables are generated from users' specified relations and home country, respectively (Burke et al., 2013; Herdağdelen et al., 2016), while for SHARE, it is what the participant declared. We control by *age* and *sex*, but we do not use the variable *education* because—as we show later—there is a high level of missing data for Facebook that contributes to producing a low correlation between Facebook data and SHARE-Internet data. For more information regarding these variables, the reader can refer to the data section and to Appendix A.

One important reason to use the logit model is that the Facebook data are aggregated counts: we do not have micro-level data. However, for the logit model, the maximum likelihood estimates and standard errors are the same if we use the individual-level outcomes, or if we aggregate and classify them according to their categorical independent variables (Agresti, 2013 [chap. 4, example 4.2.2]). In other words, we can obtain the same estimates from aggregate-level data, as if we had micro-level data. This statistical result, though being well known in statistics, has not been considered for the analysis of associations of Facebook variables so far. It has important implications for the research community that uses aggregate-level advertisement data, as it opens up new ways of understanding and analyzing this type of data with approaches that rely solely on aggregate-level data that are becoming more and more available but are still an untapped resource for statistical analyses.

For the analyses of SHARE data, we use the calibrated cross-sectional individual weights and consider the sample design. For a full explanation of the sample design and the weighting strategies, the reader can refer to Bergmann et al. (2017). The totals and proportions are calculated using the R package *survey* version 3.35–1 (Lumley, 2004); the logit models are also run using the survey package. For analyses of Facebook data, we programmed a bootstrap procedure to resample observation units by day in order to determine the standard errors. This way we can consider the variability of the Facebook data in terms of daily usage over time, without biasing the expected values of the estimated totals, proportions, and coefficients. A full explanation of the algorithms used to estimate the totals, proportions, and coefficients can be found in Appendix C.

Results

Representativeness of Facebook

In this section, we discuss the representativeness of Facebook users by comparing the structure of Facebook data against the information on internet use in SHARE. This is done by comparing both SHARE-Internet and Facebook calculated totals and calculated percentages by demographic characteristics (Table 1). But first, we would like to acknowledge that, overall, SNS users are more highly educated and more skilled at using the internet than the general population (Hargittai, 2020; Hunsaker & Hargittai, 2018); and that the Facebook penetration rates for older people tend to be smaller than 30% worldwide (Gil-Clavel & Zagheni, 2019). In the case

Table 1 Totals and percentages of population by characteristic

Variable	SHARE-Internet users		Facebook users	
	Totals (standard error)	% (standard error)	Totals (standard error)	% (standard error)
Sex				
Male	35,839,112	51	13,923,826	44
	(5,016,980)	(0.0063)	(173,164)	(0.0037)
Female	33,979,165	49	17,549,584	56
	(5,283,971)	(0.0063)	(141,902)	(0.0037)
Age				
50–64	47,152,426	68	23,447,374	74
	(6,607,756)	(0.0101)	(205,800)	(0.0027)
65+	22,665,851	32	8,026,036	26
	(3,746,581)	(0.0101)	(88,237)	(0.0027)
Education				
Below College	67,387,133	97	12,722,804	40
	(9,811,279)	(0.0029)	(146,152.01)	(0.0035)
College or above	1,206,362	2	1,072,054	3
	(333,710)	(0.0025)	(13,402)	(0.0005)
Unspecified	1,224,782	2	17,678,552	56
	(196,844)	(0.0019)	(162,924)	(0.0035)
Parent				
No	7,717,093	11	25,783,529	82
	(830,401)	(0.0079)	(213,323)	(0.0020)
Yes	62,101,184	89	5,689,880	18
	(9,546,248)	(0.0079)	(61,880)	(0.0020)
Immigrant				
No	69,060,400	99	28,559,020	91
	(10,247,107)	(0.0021)	(220,529)	(0.0009)
Yes	757,877	1	2,914,390	9
	(124,946)	(0.0021)	(21,718)	(0.0009)
Friend				
No	50,142,763	72	27,788,427	88
	(6,965,994)	(0.0101)	(217,434)	(0.0013)
Yes	19,675,514	28	3,684,982	12
	(3,377,911)	(0.0101)	(35,385)	(0.0013)

The values for SHARE were estimated using sample weights; for Facebook we bootstrapped by day. Beneath each value is its standard error

of older people in Europe, using as denominator the SHARE total populations, the percentage of Facebook users in the total population age 50 and over is 21.45% for men and 22.59% for women. Therefore, for the European 50+ population, Facebook users do not represent a random sample of the general population. However, with a

continuously increase of internet usage for the older age groups (Hunsaker & Hargittai, 2018), we can expect a continuous increase of SNS usage as well—as shown at the end of this section. Furthermore, as Table 1 shows, the percentages present important similarities, though the total number of users in Facebook is a fraction of those that use internet in SHARE.

In the case of sex, we see that, while in SHARE-Internet there are 2 percentage points more men than women, in Facebook there is a difference of 12 percentage points in favor of women. This outcome is also observed by Gil-Clavel and Zagheni (2019) in their results for Facebook users in Europe, where the median ratios of female users by country population are always greater than for men. Regarding age, we observe that in both databases there are more people in the younger age group, between 50 and 64, using internet and Facebook (two thirds and three fourths of the populations, respectively). For education, a large portion of the (self-reported) values are missing in Facebook: the unspecified category has a difference of 54 percentage points between the SHARE-Internet users and Facebook users. This difference skews the values for the other two categories, making the Facebook percentages differ from the SHARE-Internet ones. The high percentage of unspecified level of education is also observed by Ribeiro et al. (2020) in their study of the population from the United States.

When comparing the variable parent, we see that 89% of the SHARE-Internet users are parents, while in Facebook this is only 18%. This might be because Facebook likely does not identify many parents or users do not disclose their family ties, as this requires users to explicitly make the links in the SNS. In the case of immigrants, the percentages in both databases are similar, with more than 90% of the population not having this attribute. For the variable friend, we observe that in both cases less than one third of the populations can be considered as having close friends, 28% and 12% for SHARE-Internet and Facebook respectively.

Figure 1 shows the demographic distribution of Facebook and internet users by characterization as described in Eq. (1). When we consider education, the Pearson correlation between the Facebook and SHARE-Internet users is 0.44 (CI (95%): [0.18, 0.64]): this can be seen in Fig. 1a and is a consequence of the large fraction of Facebook users that do not disclose their educational history.

If we do not break down by level of education ($P_{ij}(s, a) = \sum_e P_{ij}(s, a, e)$), the percentages have a Pearson correlation of 0.77 (CI(95%): [0.45, 0.92]). Figure 1b shows that the relation between these proportions is in general very linear (the values can be found in Fig. 3 of Appendix D). The only exception is the immigrant population, where the point 50–64 male takes values of 60% in SHARE and 34% in Facebook. This results in an underrepresentation of the male 50–64 population in Facebook, while the rest of the groups are overrepresented in the SNS. We also observe that, on the one hand, for the 65 + population the values for women and men by characterization do not differ that much. On the other hand, for the 50–64 population, women are overrepresented in Facebook and men are underrepresented.

The three main highlights of this analysis are: (1) at the population level we observe that, while Facebook users are only a fraction of the total internet users, the distribution of demographic features is highly correlated across the two populations, except for the educational variable, which shows not to be a reliable measure
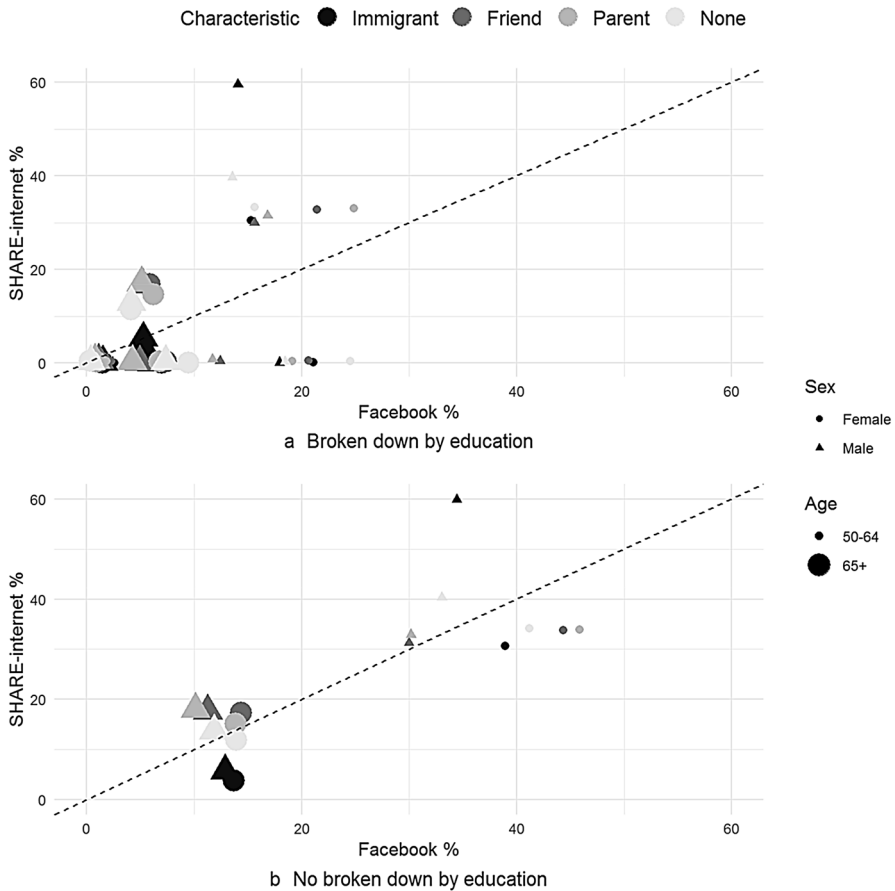
Fig. 1 Relationship between Facebook and SHARE-Internet proportions by characteristic. The red dashed line is the identity function

in the Facebook dataset; (2) although there are substantially more parents using internet than those estimated to be parents in Facebook, indicating that Facebook likely does not identify many parents or users do not disclose their family ties, the demographic characteristics of parents in Facebook are linearly correlated with the ones in the SHARE-Internet users database; (3) the male 50–64 immigrant group is highly underrepresented in Facebook, while the rest of the immigrant groups are overrepresented.

Characteristics of Close Social Networks

The original numerical results from the logit models for having close friends in SHARE and Facebook are summarized in Table 3 of Appendix E. Figure 2 shows a dot plot with bars corresponding to 95 percent confidence intervals of the estimated
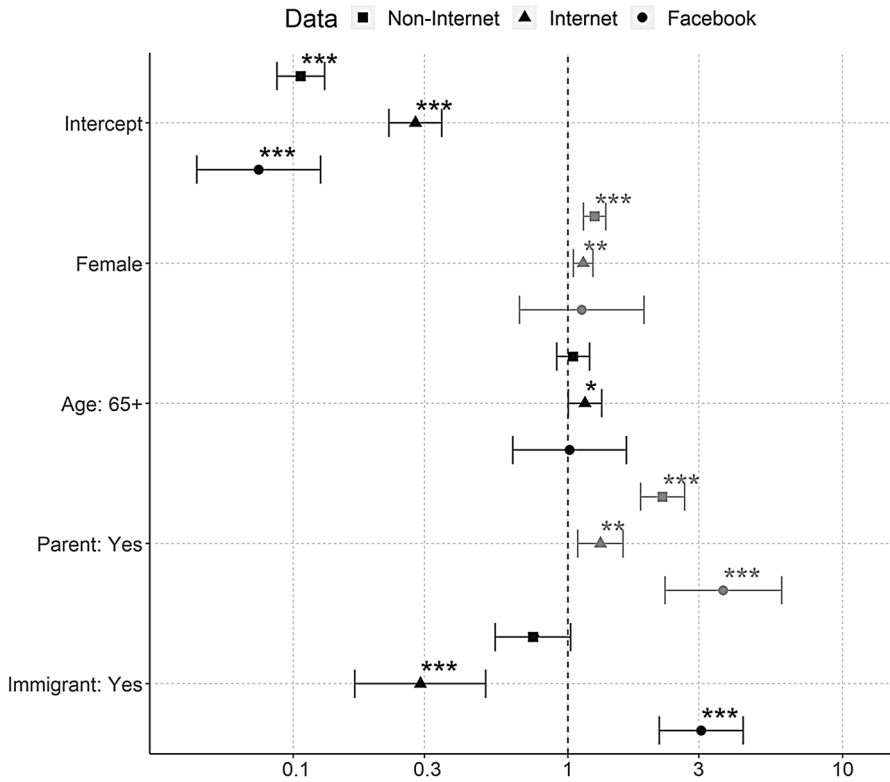
Fig. 2 Odd ratios and confidence intervals (95%) of the friend logit models. The values were estimated using the survey weights for SHARE and bootstrapped for Facebook. Significance codes: ***p value < 0.001, **p value < 0.01, *p value < 0.05. The dashed line corresponds to the one x-axis intersection. Original values in log 10 scale

Table 2 Results from the Wald test for no difference in two independent samples applied to the logit models

	Non-Internet vs. Internet	Non-Internet vs. Facebook
Intercept	− 0.9605*** (− 1.2588, − 0.6622)	0.3394 (− 0.223, 0.9018)
Sex: Female	0.0972 (− 0.0275, 0.2219)	0.1183 (− 0.4082, 0.6448)
Age:65 +	− 0.0996 (− 0.2961, 0.0969)	0.0233 (− 0.4811, 0.5277)
Parent: Yes	0.5217*** (0.257, 0.7864)	− 0.5021 (− 1.0348, 0.0306)
Immigrant: Yes	0.9451** (0.3111, 1.5791)	− 1.4094*** (− 1.8762, -0.9426)

Coefficients (log odds) with 95% confidence intervals and significance codes: ***p value < 0.001, **p value < 0.01, *p value < 0.05

odd ratios from the logistic regression and Table 2 shows the results from the Wald test (the values are also in Table E1). The baseline values represent the population of individuals between 50 and 64 that are men, not parents, and not immigrants.

The baseline probabilities of having a close friend are $0.1064/(1 + 0.1064) =$ 9.6% for non-internet users, 21.7% for internet users, and 7.04% for Facebook users (for Facebook users, this refers to having an online close friend). Being a woman increases the odds of having close friends by 22% for non-internet users, while for both internet and Facebook users, the odds increase by 10%. According to the Wald test, we cannot reject the null hypothesis that the difference between the non-internet model and both the internet and Facebook models are zero. In the case of age, we see that there is no statistical evidence that this variable is associated with having close friends either offline or online. For the internet model, we observe that the p value of the age coefficient is 0.0455, positioning the coefficient at the limit of not being significant, according to traditional definitions.

Being a parent has a positive association with having close friends. It increases the probability to $0.1064 \times 2.2107/(1 + 0.1064 \times 2.2107) = 19\%$ for non-internet users. In the online case, the probability increases by 4 and 14 percentage points for internet and Facebook users, respectively. The variable immigrant has a negative association with having close friends for internet users, decreasing the probabilities to 7.5%, while for non-internet users, we cannot discard that there is no association. For Facebook users the association is positive, increasing the probability from 7.2% to 18.7%.

In summary, the main results shown in this section are the following ones: (1) being a woman has very similar effects on the probability of having close friends regardless of whether she uses internet or not, and a similar size effect is estimated for the probability of having close friends online; (2) age among older adults does not play a central role in determining the likelihood of having close friends either online or offline; (3) being a parent has always a positive association with having close friends; and (4) while being a migrant is not associated with having close friends for those offline, it is negatively associated for those that use internet and positively associated for Facebook users.

Conclusion and Discussion

Research on the use of Information and Communication Technologies by older adults points to an increase in their access to social capital (Jung & Sundar, 2016; Neves and Barbara 2013). This might be because these technologies facilitate older adults' communication with their families and friends regardless of geographical distance (Neves et al., 2018; Quan-Haase et al., 2018). In particular, older migrants might be the ones that benefit the most, given that they tend to have more links with people living abroad, and thus far away, compared with natives (Baldassar et al., 2016; Näre et al., 2017). This work offers three main contributions. First, we analyze the representativeness of older people in Facebook. Second, we test whether the internet and Facebook associations are statistically different from the offline ones.

Finally, we analyze, at the population level, which older people are more likely to have close friends (or online ties) depending on whether they are internet users, Facebook users, or offline.

To study the representativeness of Facebook data for aging research, we compare the demographic features of the SHARE respondents who use internet (SHARE-Internet users) with those of Facebook users. We find that the demographic structure of the Facebook data is highly correlated with the structure of the SHARE-Internet users when we do not break down the sample by level of education. This is because a large fraction of Facebook users does not disclose their educational history, resulting in high percentages of users with unspecified level of education, thus, making the Facebook variable education not comparable with SHARE-Internet.

Concerning information about migration background, the structure of the data differs between the two sources. This might be a consequence of the kind of migration that SHARE and Facebook capture. On the one hand, SHARE respondents not born in the country of interview tend to have lived there for more than 40 years (Bordone & De Valk, 2016) possibly not having strong connections with the country of origin any longer. On the other hand, Ciobanu et al. (2017) notice that most international retirement migrants do not learn the host country's language. Therefore, given the restrictions that SHARE imposes on the people considered for interview, the survey might not capture retirement migrants, while Facebook might. A second kind of migration that Facebook might be capturing is the zero-generation, parents of migrant children who follow their adult children in migration or engage in back-and-forth mobility as a medium for inter-generational support (Ciobanu et al., 2017).

Regarding whether older people who are internet users or Facebook users (as a proxy for SNS usage) are more likely to have close friends or online close friends than those who are offline, we observe that being a woman is positively associated with having close friends both online and offline. However, the difference between the non-internet and both the internet and Facebook coefficients is not statistically significant, which can be interpreted as the differences in the means being zero. We find no statistical evidence that age plays a major role among older adults in having close friends neither offline nor online, whereas being a parent has a positive association in both cases. Our results corroborate the findings from McPherson et al. (2006) for the American population. In general, people lose or cut contact with acquaintances as they get older but maintain strong ties with their nuclear family and women have a slight advantage over men in maintaining their friends.

In the case of migrants, our analysis shows that for the non-internet users being a migrant does not play a role on having close friends, whereas for internet users the association is negative. This might be related to a selection effect, where those who have fewer friends are more likely to use internet. However, more research has to be done to study the relationship between internet use and friendships. In the case of Facebook, the association between having close friends online and being an immigrant is positive, which has also been reported in qualitative analyses (Baldassar et al., 2016; Bates & Komito, 2012; Komito, 2011). This suggests that older migrants may be more likely to use SNS to maintain social relationships. Interaction online can partially compensate for the lower level of close friends offline and could

be the result of a selection process whereby having fewer friends offline might lead migrants to establish or maintain digital friendships.

Our work has important limitations that we should acknowledge. First, we assume that Facebook is a proxy for use of all SNS. This is not necessarily the case. However, Facebook is currently the biggest SNS worldwide; therefore, a large section of social network site users are Facebook users. So, we can still think about our results as representative of a large number of SNS users. Second, we use data available from the Facebook Marketing API, which has limitations because it was not produced for research. For this paper, we inferred the definitions of the Facebook variables from publications that come from the Facebook Data Science Team. We acknowledge that measurement error is likely present in the data. For example, there is an increased likelihood that there are unidentified parents among the reference group (classified as non-parents). This misclassification will weaken the difference between those classified as parents and non-parents, i.e., move our effect estimates towards the null. As a consequence, our conclusions are conservative. Despite this limitation, we were able to find differences between groups. We acknowledge that this is a weakness in our work that cannot be addressed directly, but future work can build on this further by testing our results using different datasets. Further research also needs to be done in order to have a better understanding of how different types of measurements were operationalized. This will be a continuous process that involves two avenues that we are already pursuing: on the one hand, we are developing research partnerships with the Facebook Data Science Team that has access to raw data, in addition to aggregate estimates. On the other hand, we are working on developing surveys of Facebook users that can give us more information about the biases in the data and the reliability of different types of measures of socio-demographic characteristics (Grow et al., 2021; Grow et al., 2020).

Third, we focus on countries that are represented in SHARE in order to anchor our analysis to a probabilistic survey. However, further research with Facebook data can go beyond Europe to assess how social networking sites affect access to social capital across a broader range of geographic settings, including at the subnational level. Even though there are clear limitations when working with digital trace data, we believe that there is value in combining analyses that include both probabilistic surveys and passively collected information. We hope that this article, beyond providing substantive results, also contributes to a methodological discussion on how to best use increasingly available digital trace data to complement surveys.

We expect future research to further develop the combined use of digital traces and survey data for advancing our understanding of the size and type of social networks over the life course. Our use of digital traces is limited to Facebook. While Facebook is the largest social media platform, social interaction happens also in other spaces in internet. Feehan and Cobb (2019) developed a methodology to assess the characteristics of non-Facebook users when interviewing a sample of Facebook users. More specifically, they used a network reporting approach that relies on the idea of asking respondents sampled via Facebook to report about a specific feature (e.g., internet adoption) among other people they are connected to in their everyday offline personal networks. As running surveys where participants are recruited via social media platforms like Facebook is becoming more and more feasible and

relevant (Grow et al., 2020; Kühne & Zindel, 2020), and methods to assess the biases of these approaches are rapidly evolving (Grow et al., 2021; Sances, 2019), we expect that the combination of surveys and digital traces will play an increasingly central role in expanding our knowledge of social networks.

In summary, in this article, we studied the association between older people's characteristics and the likelihood of having close friends offline and online. Our statistical analysis concluded that being online has an important differential effect for the population of migrants. In particular, we estimated a positive association between being a migrant and having close friends online for older people using Facebook. Previous research has highlighted the health benefits that the use of Information and Communication Technologies could bring to older people (Rios et al., 2019; Sum et al., 2008), as these technologies ease older people's access to social capital (Neves and Barbara 2013). In this work, we show that, among older people, the ones that seem to benefit the most from these technologies are migrants. While more research has to be done to understand the potential causal mechanisms behind what we observed, our article also made a methodological contribution to the study of online relationships by showing how classic regression models can be leveraged when using freely available aggregate-level data from advertisement platforms of major social media companies.

Appendix A

SHARE Variables

In parentheses are the original SHARE names of the variables.

- age_{SHARE}: Transformation of the SHARE variable year of birth (dn003_) to a categorical variable with the values 50–64 or 65+.
- $education_{SHARE}$: The highest school degree obtained according to the ISCED classification (isced1997_r):
  - Below College: either ISCED-97 code 1, ISCED-97 code 2, ISCED-97 code 3, ISCED-97 code 4, or ISCED-97 code 5.
  - College or Above: ISCED-97 code 6.
  - Unspecified: either Refusal, Don't know, None, Other, or Still in school.
- $parent_{SHARE}$: Transformation of the number of children (ch001_) into a dummy variable with 1 meaning the person has at least one child and 0 otherwise.
- $immigrant_{SHARE}$: It is a dummy variable with 1 meaning the person was not born in the country of interview (dn004_) and 0 otherwise.
- $friend_{SHARE}$: Transformation of the variable network closeness (sn009_X) into a dummy variable with 1 when the person declared to have at least one person with whom she or he feels *Somewhat close, Very close* or *Extremely close* and 0 otherwise.

- *internet$_{SHARE}$*: Dummy variable of whether the person used internet during the past 7 days, either for e-mailing, searching for information, making purchases, or for any other purpose at least once (it004_). With one meaning *yes* and zero *no*.

Facebook Variables

In parentheses are the original Facebook names[5] of the variables.

- country: geographical targeting field from country, region, city or zip (geo_locations). In our case any of the 17 countries show in the Appendix A.
- *sex$_{FB}$*: User declared gender either male (0) or female (1).
- *age$_{FB}$*: User declared age. Either between ages 50 and 64 (0) or 65 or above (1).
- *education$_{FB}$*: User declared education (education_statuses). It is a categorical variable with the next categories- using Facebook nomenclature[6]:
    - Below College: either HIGH_SCHOOL, UNDERGRAD, HIGH_SCHOOL_ GRAD, or SOME_HIGH_SCHOOL.
    - College or Above: either SOME_COLLEGE, ASSOCIATE_DEGREE, IN_ GRAD_SCHOOL, SOME_GRAD_SCHOOL, MASTER_DEGREE, PRO-FESSIONAL_DEGREE, or DOCTORATE_DEGREE.
    - Unspecified: UNSPECIFIED.
- *parent$_{FB}$*: People who are parents (Parents (All)), coded as 1 = yes and 0 = no.
- *immigrant$_{FB}$*: People living outside their home country (Expats (All)), coded as 1 = yes and 0 = no.
- *friend$_{FB}$*: Close friends of people with a birthday in 7–30 days (Close friends of people with birthdays in a month), coded as 1 = yes and 0 = no.

Appendix B

List of countries that constitute the bases:

Austria	Italy
Belgium	Luxembourg
Croatia	Poland
Czechia	Portugal
Denmark	Slovenia
Estonia	Spain
France	Sweden
Germany	Switzerland
Greece	

---

[5] https://developers.facebook.com/docs/marketing-api/audiences/reference/basic-targeting. Accessed 30 September 2020.

[6] https://developers.facebook.com/docs/marketing-api/audiences/reference/advanced-targeting#education_and_workplace. Accessed 30 September 2020.

Appendix C

In this section, we explain the bootstrap procedure that we follow to estimate totals, percentages, and coefficients. First, let us stablish some notation, $\#Users_i(\dots)$ is the number Facebook Daily Active Users that the Facebook Marketing API returned for day i and that have the following characteristics:$country \in$\{Any of the 17 countries of Appendix B\} $sex \in$\{Female, Male\}; $age \in$\{50–64, 65+\}; $education \in$\{Unspecified, Below College, College or Above\}.; $immigrant \in$\{0,1\}, $friend \in$\{0,1\}, $parent \in$\{0,1\}.

$$\#Users_i(country, sex, age, education, immigrant, friend, parent). \tag{3}$$

This way the Facebook daily active users that are women, live in Spain, are older than 65, with an education below college, and are mothers, migrants, and that have close friends in day 14 is

$$\#Users_{14}(Spain, female, 65+, belowcollege, 1,1,1).$$

Then our database consists of 1632 characteristics combinations and 31 possible values, because these values can vary each day.

The bootstrap procedure to obtain the totals and ratios consists of two steps. First, we create n (in this case n was 2500) number of subsamples. Second, we calculate the means, ratios, and standard errors. For the first step, for each of the 1632 characteristics combinations, we randomly select one of the 31 values. This way one subsample would consist of 1632 characteristics combinations, but the values come from different days of observation units per characteristic. To be clearer, in one subsample, we can have $\#Users_{14}$(Spain, female, 65+, bellow-college,1,1,1), $\#Users_3$(Spain, male, 65+, bellow-college,1,1,1) that $\#Users_{30}$(Italy, female, 50–64, bellow-college,0,0,0), etc.

For the second step, for each category (Cat) of the variables (Var) sex, age, education, immigrant, friends, and parent, we calculate the totals and percentages following Eqs. (4) and (5), respectively.
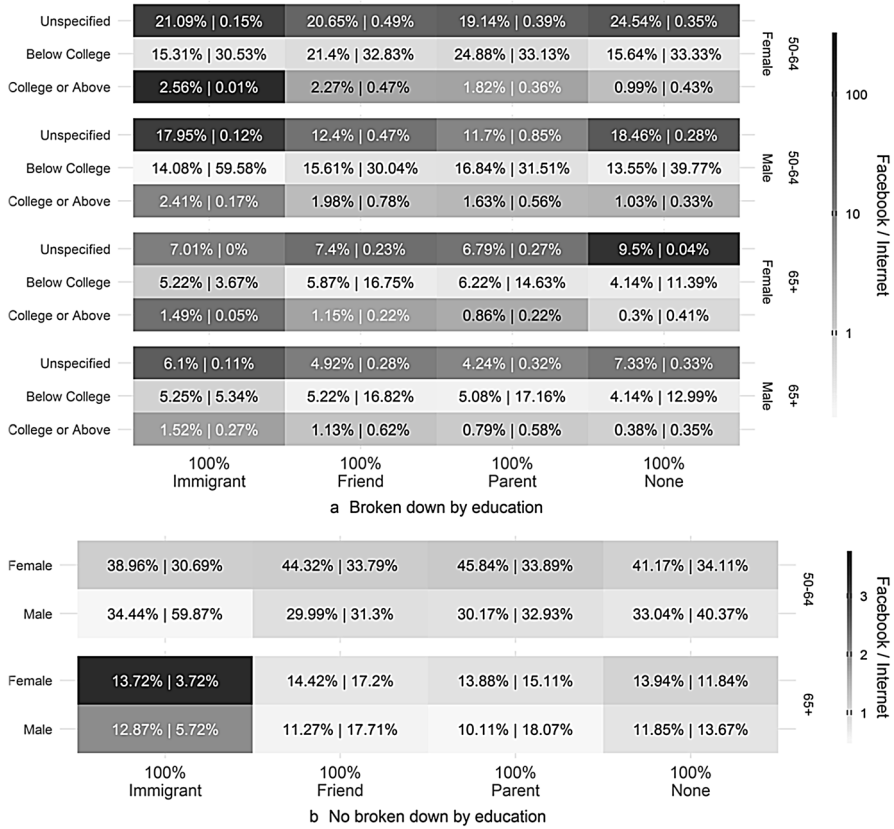
$$Total_{Cat} = \sum \#Users(Var == Cat), \tag{4}$$

$$Percentage_{Cat} = \frac{Total_{Cat}}{\sum_{i \in Var} Total_i}. \tag{5}$$

The bootstrap procedure to estimate coefficients also consists of two steps. The first step is the same as for the totals and ratios. For the second step, we calculate the coefficients of the generalize linear model of Eq. (2) of the methodology for each subsample of step one. Then we obtain the mean, standard deviation, 95% confidence intervals, and p values using the coefficients from all subsamples. In R, the glm function would look like Code C1.

$glm(friend \sim sex + age + parent + immigrant,$
$data = subsample_i, weights = \#Users, family = binomial() \dots CodeC1).$

Appendix D

See Fig. 3.

	100% Immigrant	100% Friend	100% Parent	100% None		
Unspecified	21.09% | 0.15%	20.65% | 0.49%	19.14% | 0.39%	24.54% | 0.35%	Female	50-64
Below College	15.31% | 30.53%	21.4% | 32.83%	24.88% | 33.13%	15.64% | 33.33%		
College or Above	2.56% | 0.01%	2.27% | 0.47%	1.82% | 0.36%	0.99% | 0.43%		
Unspecified	17.95% | 0.12%	12.4% | 0.47%	11.7% | 0.85%	18.46% | 0.28%	Male	50-64
Below College	14.08% | 59.58%	15.61% | 30.04%	16.84% | 31.51%	13.55% | 39.77%		
College or Above	2.41% | 0.17%	1.98% | 0.78%	1.63% | 0.56%	1.03% | 0.33%		
Unspecified	7.01% | 0%	7.4% | 0.23%	6.79% | 0.27%	9.5% | 0.04%	Female	65+
Below College	5.22% | 3.67%	5.87% | 16.75%	6.22% | 14.63%	4.14% | 11.39%		
College or Above	1.49% | 0.05%	1.15% | 0.22%	0.86% | 0.22%	0.3% | 0.41%		
Unspecified	6.1% | 0.11%	4.92% | 0.28%	4.24% | 0.32%	7.33% | 0.33%	Male	65+
Below College	5.25% | 5.34%	5.22% | 16.82%	5.08% | 17.16%	4.14% | 12.99%		
College or Above	1.52% | 0.27%	1.13% | 0.62%	0.79% | 0.58%	0.38% | 0.35%		

Facebook / Internet (100, 10, 1)

a Broken down by education

	100% Immigrant	100% Friend	100% Parent	100% None		
Female	38.96% | 30.69%	44.32% | 33.79%	45.84% | 33.89%	41.17% | 34.11%	50-64	
Male	34.44% | 59.87%	29.99% | 31.3%	30.17% | 32.93%	33.04% | 40.37%		
Female	13.72% | 3.72%	14.42% | 17.2%	13.88% | 15.11%	13.94% | 11.84%	65+	
Male	12.87% | 5.72%	11.27% | 17.71%	10.11% | 18.07%	11.85% | 13.67%		

Facebook / Internet (3, 2, 1)

b No broken down by education

Fig. 3 Proportions of Facebook and SHARE-Internet users by characterization and demographic characteristics. In the cells, the left value corresponds to Facebook and the right value to SHARE-Internet. The color shows the ratio of Facebook percentage by Internet percentage

Appendix E

See Table 3.

Table 3 Log odds and confidence intervals of the friend logit models and Wald Test results

	Log odds			Wald test	
	No Internet (NI)	Internet (I)	Facebook (FB)	NI–I	NI–FB
Intercept	− 2.2402***	− 1.2797***	− 2.5796***	− 0.9605***	0.3394
	(− 2.4407, − 2.0397)	(− 1.5007, − 1.0587)	(− 3.1050, − 2.0542)	(− 1.2588, − 0.6622)	(− 0.223, 0.9018)
Sex: Female	0.2247***	0.1275**	0.1063	0.0972	0.1183
	(0.1313, 0.3182)	(0.0449, 0.2100)	(− 0.4118, 0.6245)	(− 0.0275, 0.2219)	(− 0.4082, 0.6448)
Age: 65 +	0.0433	0.1429*	0.0200	− 0.0996	0.0233
	(− 0.0947, 0.1813)	(0.0029, 0.2828)	(− 0.4651, 0.5052)	(− 0.2961, 0.0969)	(− 0.4811, 0.5277)
Parent: Yes	0.7933***	0.2716**	1.2953***	0.5217***	− 0.5021
	(0.6091, 0.9775)	(0.0815, 0.4617)	(0.7955, 1.7952)	(0.257, 0.7864)	(− 1.0348, 0.0306)
Immigrant: Yes	− 0.2930	− 1.2381***	1.1163***	0.9451**	− 1.4094***
	(− 0.6093, 0.0234)	(− 1.7874, − 0.6888)	(0.7730, 1.4597)	(0.3111, 1.5791)	(− 1.8762, − 0.9426)

Coefficients (Log Odds) with Confidence Intervals (95%) and Significance Codes: ***p value < 0.001, **p value < 0.01, *p value < 0.05

Declarations

Conflicts of interest The authors declare that they have no conflict of interests.

Data Availability SHARE data are available free of charge at http://www.share-project.org. Regarding the Facebook data, we will make the data available via email to individual researchers for the purposes of reproducing the results. We will not post the data on a freely accessible webpage as it goes against the Facebook's terms and conditions.

References

Agresti, A. (2013). *Categorical data analysis. A Wiley-Interscience Publication*. Wiley.

Alburez-Gutierrez, D., Zagheni, E., Aref, S., Gil-Clavel, S., Grow, A., & Negraia, D. V. (2019). Demography in the Digital Era: New Data Sources for Population Research. In *Smart Statistics for Smart Applications: Book of Short Papers SIS2019, Arbia, Giuseppe, et al*. Pearson. https://osf.io/24jp7.

Backstrom, L., Bakshy, E., Kleinberg, J., Lento, T. M., & Rosenn, I. (2011). Center of attention: How Facebook users allocate attention across friends. In *2011, 8. Proc. 5th International AAAI Conference on Weblogs; Social Media*.

Baldassar, L., Nedelcu, M., Merla, L., & Wilding, R. (2016). ICT-based co-presence in transnational families and communities: Challenging the premise of face-to-face proximity in sustaining relationships. *Global Networks, 16*(2), 133–144. https://doi.org/10.1111/glob.12108

Ballantyne, A., Trenwith, L., Zubrinich, S., & Corlis, M. (2010). 'I feel less lonely': What older people say about participating in a social networking website. *Quality in Ageing and Older Adults, 11*(3), 25–35. https://doi.org/10.5042/qiaoa.2010.0526

Bates, J., & Komito, L. (2012). Migration, community and social media. In G. Boucher, A. Grindsted, & T. L. Vicente (Eds.), *Transnationalism in the global city. Maior Series 6.* University of Deusto.

Beauchemin, C., & González-Ferrier, A. (2011). Sampling international migrants with origin-based snowballing method: New evidence on biases and limitations. *Demographic Research, 25*(July), 103–134. https://doi.org/10.4054/DemRes.2011.25.3

Bergmann, M., De Luca, G., & Scherpenzeel, A. (2017). Sampling design and eighting strategies in SHARE Wave 6. In R. Ramya (Ed.), *SHARE Wave 6: Panel innovations and collecting dried blood spots.* MEA, Max Planck Institute for Social Law; Social Policy.

Bordone, V., & de Valk, H. A. G. (2016). Intergenerational support among migrant families in Europe. *European Journal of Ageing, 13*(3), 259–270. https://doi.org/10.1007/s10433-016-0363-6

Bourdieu, P. (1986). *The forms of social capital*.

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., & Zuber, S. (2013). Data resource profile: The survey of health, ageing and retirement in Europe (SHARE). *International Journal of Epidemiology, 42*(4), 992–1001. https://doi.org/10.1093/ije/dyt088

Brame, R., Paternoster, R., MazeroIIe, P., & Piquero, A. (1998). Testing for the equality of maximum-likelihood regression coefficients between two independent equations. *Journal of Quantitative Criminology, 14*(3), 17.

Burke, M., Adamic, L. A., & Marciniak, K. (2013). Families on Facebook. *In Seventh International AAAI Conference on Weblogs and Social Media, 10*.

Ciobanu, R. O., Fokkema, T., & Nedelcu, M. (2017). Ageing as a migrant: Vulnerabilities, agency and policy implications. *Journal of Ethnic and Migration Studies, 43*(2), 164–181. https://doi.org/10.1080/1369183X.2016.1238903

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94*, 27.

Dubois, A., Zagheni, E., Garimella, K., & Weber, I. (2018). *Studying migrant ***assimilation through Facebook interests*. [Cs], January.

Edelmann, A., Wolff, T., Montagne, D., & Bail, C. A. (2020). Computational social science and sociology. *Annual Review of Sociology, 46*(1), 61–81. https://doi.org/10.1146/annurev-soc-121919-054621

Eurostat. (2018). *ICT Usage in Households and by Individuals (Isoc_i). Data base*. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc_ci_ifp_iu&lang=en

Facebook Inc. (2019). *Facebook reports third quarter 2019 results*. Retrieved September 24, 2020, from https://investor.fb.com/investor-news/press-release-details/2019/Facebook-Reports-Third-Quarter-2019-Results/default.aspx

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

Feehan, D. M., & Cobb, C. (2019). Using an online sample to estimate the size of an offline population. *Demography, 56*(6), 2377–2392. https://doi.org/10.1007/s13524-019-00840-z

Francis, J., Ball, C., Kadylak, T., & Cotten, S. (2019). Aging in the digital age: Conceptualizing technology adoption and digital inequalities. In B. B. Neves & F. Vetere (Eds.), *Ageing and digital technology: Designing and evaluating emerging technologies for older adults.* Springer.

Friemel, T. N. (2016). The digital divide has grown old: Determinants of a digital divide among seniors. *New Media & Society, 18*(2), 313–331. https://doi.org/10.1177/1461444814538648

Gil-Clavel, S., & Zagheni, E. (2019). Demographic differentials in facebook usage around the world. In *Proceedings of the International AAAI Conference on Web and Social Media 13*(01), 647–650. https://aaai.org/ojs/index.php/ICWSM/article/view/3263

Goodreau, S. M., Kitts, J. A., & Morris, M. (2009). Birds of a feather, or friend of a friend?: Using exponential random graph models to investigate adolescent social networks. *Demography, 46*(1), 103–125. https://doi.org/10.1353/dem.0.0045

Grow, A., Perrotta, D., Del Fava, E., Cimentada, J., Rampazzo, F., Gil-Clavel, S., & Zagheni, E. (2020). Addressing public health emergencies via Facebook surveys: Advantages, challenges, and practical considerations. *Journal of Medical Internet Research, 22*(12), 12. https://doi.org/10.2196/20653

Grow, A., Perrotta, D., Del Fava, E., Cimentada, J., Rampazzo, F., Gil-Clavel, S., Zagheni, E., Flores, R. D., Ventura, I., & Weber, I. (2021). *How reliable is Facebook's advertising data for use in social*

science research? Insights from a cross-national online survey. WP-2021006. Max Planck Institute for Demographic Research.

Hargittai, E. (2020). Potential biases in big data: Omitted voices on social media. *Social Science Computer Review, 38*(1), 10–24. https://doi.org/10.1177/0894439318788322

Herdağdelen, A., State, B., Adamic, L., & Mason, W. (2016). The social ties of immigrant communities in the United States. In *Proceedings of the 8th ACM Conference on Web Science - WebSci '16*, (78–84). ACM Press. https://doi.org/10.1145/2908131.2908163

Hunsaker, A., & Hargittai, E. (2018). A review of internet use among older adults. *New Media & Society, 20*(10), 3937–3954. https://doi.org/10.1177/1461444818787348

Jung, E. H., & Shyam Sundar, S. (2016). Senior citizens on Facebook: How do they interact and why? *Computers in Human Behavior, 61*(August), 27–35. https://doi.org/10.1016/j.chb.2016.02.080

Komito, L. (2011). Social media and migration: Virtual community 2.0. *Journal of the American Society for Information Science and Technology, 62*(6), 1075–1086. https://doi.org/10.1002/asi.21517

Kühne, S., & Zindel, Z. (2020). Using Facebook and Instagram to recruit web survey participants: A step-by-step guide and application. *Survey Methods: Insights from the Field*. https://doi.org/10.13094/SMIF-2020-00017

Lazer, D., & Radford, J. (2017). Data Ex Machina: Introduction to big data. *Annual Review of Sociology, 43*(1), 19–39. https://doi.org/10.1146/annurev-soc-060116-053457

Lee, B., Chen, Y., & Hewitt, L. (2011). Age differences in constraints encountered by seniors in their use of computers and the internet. *Computers in Human Behavior, 27*(3), 1231–1237. https://doi.org/10.1016/j.chb.2011.01.003

Lee, K.-T., Noh, M.-J., & Koo, D.-M. (2013). Lonely people are no longer lonely on social networking sites: The mediating role of self-disclosure and social support. *Cyberpsychology, Behavior, and Social Networking, 16*(6), 413–418. https://doi.org/10.1089/cyber.2012.0553

Lin, N. (1999). Building a network theory of social capital. *Connections, 22*, 24.

Lumley, T. (2004). Survey: Analysis of complex survey samples. *Journal of Statistical Software, 9*(1), 19.

Lumley, T. S. (2010). *Complex surveys: A guide to analysis using R*. Wiley.

Maya Jariego, I. (2018). Why name generators with a fixed number of alters may be a pragmatic option for personal network analysis. *American Journal of Community Psychology, 62*(1–2), 233–238. https://doi.org/10.1002/ajcp.12271

McPherson, M., Smith-Lovin, L., & Brashears, M. E. (2006). Social isolation in America: Changes in core discussion networks over two decades. *American Sociological Review, 71*(3), 23.

Merluzzi, J., & Burt, R. S. (2013). How many names are enough? Identifying network effects with the least set of listed contacts. *Social Networks, 35*(3), 331–337. https://doi.org/10.1016/j.socnet.2013.03.004

Näre, L., Walsh, K., & Baldassar, L. (2017). Ageing in transnational contexts: Transforming everyday practices and identities in later life. *Identities, 24*(5), 515–523. https://doi.org/10.1080/1070289X.2017.1346986

Neal, Z. P., & Neal, J. W. (2017). Network analysis in community psychology: Looking back, looking forward. *American Journal of Community Psychology, 60*(1–2), 279–295. https://doi.org/10.1002/ajcp.12158

Neves, B. B. (2013). Social capital and internet use: The irrelevant, the bad, and the good: social capital and internet use. *Sociology Compass, 7*(8), 599–611. https://doi.org/10.1111/soc4.12059

Neves, B. B. (2015). Does the Internet matter for strong ties? Bonding social capital, internet use, and age-based inequality. *International Review of Sociology, 25*(3), 415–433. https://doi.org/10.1080/03906701.2015.1050307

Neves, B. B., Baecker, R., Carvalho, D. D., Sanders, A., & Casimiro, C. (2018). Cross-disciplinary research methods to study technology use, family, and life course dynamics: Lessons from an action research project on social isolation and loneliness in later life. *Connecting families* (pp. 113–132). Bristol University Press.

Peng, S., et al. (2018). Use of communication technology to maintain intergenerational contact: toward an understanding of 'digital solidarity'. *Connecting Families* 159–180.

Putnam, R. D. (1993). The prosperous community: Social capital and public life. *The American Prospect, 13*, 11.

Quan-Haase, A., Wang, H., Wellman, B., & Zhang, A. (2018). Weaving family connections on and offline: The turn to networked individualism. *Connecting Families: Information & Communication Technologies, Generations, and the Life Course* 57–77.

Ribeiro, F. N., Benevenuto, F., & Zagheni, E. (2020). How Biased is the population of facebook users? Comparing the demographics of facebook users with census data to generate correction factors. In *12th ACM Conference on Web Science* (325–334). ACM. https://doi.org/10.1145/3394231.3397923

Rios, J. S., Wohn, D. Y., & Lee, Y.-H. (2019). Effect of Internet Literacy in understanding older adults' social capital and expected internet support. *Communication Research Report, 36*(2), 93–102.

Sances, M. W. (2019). Missing the target? Using surveys to validate social media Ad targeting. *Political Science Research and Methods, 9*(1), 215–222. https://doi.org/10.1017/psrm.2018.68

Schwartz, E., Litwin, H., & Kotte, M. (2017). Measuring social network and network changes in SHARE Wave 6. *SHARE Wave 6: Panel innovations and collecting dried blood spots.* MEA, Max Planck Institute for Social Law; Social Policy.

Sims, T., Reed, A. E., & Carr, D. C. (2016). Information and communication technology use is related to higher well-being among the oldest-old. *The Journals of Gerontology Series B*. https://doi.org/10.1093/geronb/gbw130

Stewart, I., Flores, R. D., Riffe, T., Weber, I., & Zagheni, E. (2019). Rock, Rap, or Reggaeton?: Assessing Mexican Immigrants' Cultural Assimilation Using Facebook Data. In *The World Wide Web Conference* (3258–64). ACM. https://doi.org/10.1145/3308558.3313409

Suitor, J. J., Gilligan, M., & Pillemer, K. (2016). Chapter 10: Stability, change, and complexity in later-life families. In L. K. George & K. F. Ferraro (Eds.), *Handbook of aging and the social sciences* (8th ed., pp. 205–226). Academic Press.

Sum, S., Mark, R., Mathews, M. P., & Hughes, I. (2008). Internet technology and social capital: How the Internet affects seniors' social capital and wellbeing. *Journal of Computer-Mediated Communication, 14*(1), 202–220. https://doi.org/10.1111/j.1083-6101.2008.01437.x

Vitak, J. (2014). Unpacking social media's role in resource provision: Variations across relational and communicative properties. *Societies, 4*(4), 561–586. https://doi.org/10.3390/soc4040561

Warnes, A. M., & Williams, A. (2006). Older migrants in Europe: A new focus for migration studies. *Journal of Ethnic and Migration Studies, 32*(8), 1257–1281. https://doi.org/10.1080/13691830600927617

Zagheni, E., Weber, I., & Gummadi, K. (2017). Leveraging Facebook's advertising platform to monitor stocks of migrants: Leveraging Facebook's advertising Platform. *Population and Development Review, 43*(4), 721–734. https://doi.org/10.1111/padr.12102

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.