

# UC Irvine

## UC Irvine Previously Published Works

### Title

Measuring social integration and tie strength with smartphone and survey data

### Permalink

<https://escholarship.org/uc/item/0122c98p>

### Journal

PLOS ONE, 13(8)

### ISSN

1932-6203

### Authors

Dissing, Agnete S  
Lakon, Cynthia M  
Gerds, Thomas A  
[et al.](#)

### Publication Date

2018

### DOI

10.1371/journal.pone.0200678

Peer reviewed

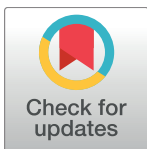
RESEARCH ARTICLE

# Measuring social integration and tie strength with smartphone and survey data

Agnete S. Dissing<sup>1\*</sup>, Cynthia M. Lakon<sup>2</sup>, Thomas A. Gerds<sup>3</sup>, Naja H. Rod<sup>4,5</sup>, Rikke Lund<sup>1,6</sup>

**1** Section of Social Medicine, Department of Public Health, University of Copenhagen, Copenhagen K, Denmark, **2** Program in Public Health, University of California Irvine, Irvine, CA, United States of America, **3** Section of Biostatistics, Department of Public Health, University of Copenhagen, Copenhagen K, Denmark, **4** Section of Epidemiology, Department of Public Health, University of Copenhagen, Copenhagen K, Denmark, **5** Copenhagen Stress Research Centre, Copenhagen, Denmark, **6** Center for Healthy Aging, Faculty of Health Sciences, University of Copenhagen, Copenhagen N, Denmark

\* [nedi@sund.ku.dk](mailto:nedi@sund.ku.dk)



## Abstract

Recordings of smartphone use for contacts are increasingly being used as alternative or supplementary measurement methods for social interactions and social relations in the health sciences. Less work has been done to understand how these measures compare with widely used survey-based information. Using data from the Copenhagen Network Study, we investigated whether derived survey and smartphone measures on two widely studied concepts; Social integration and Tie strength were associated. The study population included 737 college students (mean age 21.6 years, Standard deviation: 2.6), who were followed with surveys and continuous recordings of smartphone usage over a one-month period. We derived self-reported and smartphone measures of social integration (social role diversity, social network size), and tie strength (contact frequency, duration and tie reciprocity). Logistic regression models were used to assess the associations between smartphone derived and self-reported measures adjusting for gender, age and co-habitation. Larger call and text message networks were associated with having a high self-reported social role diversity, and a high self-reported social contact frequency was likewise associated with having both frequent call and text message interactions, longer call duration and a higher level of reciprocity in call and text message communication. Self-reported aspects of social relations and smartphone measures of social interactions have considerable overlap supporting a measurement of similar underlying concepts.

## OPEN ACCESS

**Citation:** Dissing AS, Lakon CM, Gerds TA, Rod NH, Lund R (2018) Measuring social integration and tie strength with smartphone and survey data. PLoS ONE 13(8): e0200678. <https://doi.org/10.1371/journal.pone.0200678>

**Editor:** Boris Podobnik, University of Rijeka, CROATIA

**Received:** February 17, 2017

**Accepted:** May 23, 2018

**Published:** August 23, 2018

**Copyright:** © 2018 Dissing et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data are from Copenhagen Networks study which has been approved by the Danish Data Protection Agency (DDPA) Journal nr 2012-41- 0664. DDPA is the relevant legal entity in Denmark. The full data set contains personally identifiable telecommunication patterns and survey data. According to the Act on Processing of Personal Data, such data cannot be made available in the public domain. The authors confirm that the data is available upon request to all interested researchers under conditions stipulated by the DDPA. Data inquiries should be addressed

## Introduction

Social relations are important to human health. Both structural aspects such as network size and contact frequency as well as functional aspects such as social support has been established as important determinants of human health and well-being during the last decades [1–5]. Most of this evidence is based on self-reports from surveys, however, alternative ways of measuring social relations are emerging. Over the last decade, smartphones have become increasingly available and they provide a previously unthinkable framework for gaining detailed

to the Social Fabric steering committee (<http://sodas.ku.dk>), to be reached at [ddl@econ.ku.dk](mailto:ddl@econ.ku.dk).

**Funding:** The Copenhagen Social Network Study was made possible by an interdisciplinary University of Copenhagen 2016 grant, Social Fabric (PI David Dreyer Lassen, co-PI Sune Lehman). This grant has funded purchase of the smartphones, as well as technical personnel. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

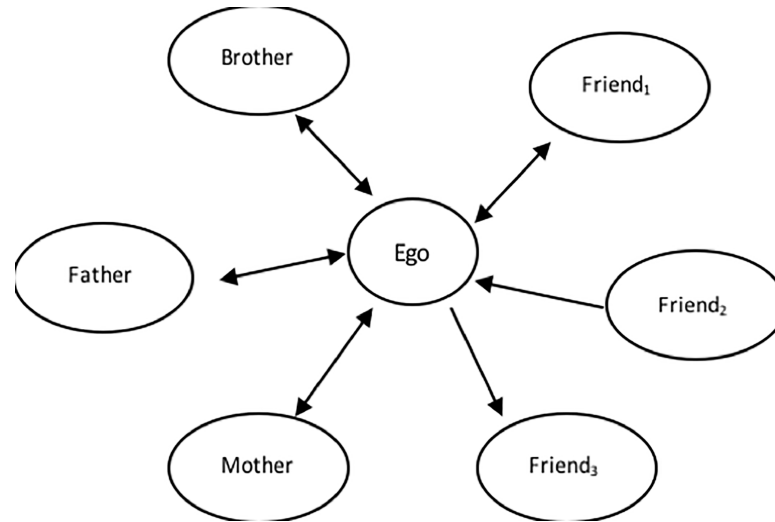
**Competing interests:** The authors have declared that no competing interests exist.

insights into human social interaction. Phone calls, online comments, GPS location and Wi-Fi-login are automatically recorded [6, 7]. These kinds of ‘big data’ provide fine grained information on human social interactions over time and place [7–12], and are increasingly being used to study social relationships in relation to health [8, 11–15], but also in relation to other areas such as transportation [16], political opinions [17], economic opportunities [18], and information spread [19]. Despite the increasing use of these data, less work has been carried out to evaluate how smartphone measures of social relations compare with more widely-used survey-based methods. Investigating whether survey and smartphone measures capture underlying theoretical concepts is important in order to further advance the use of smartphone measures of social relations.

Smartphone data reflect human social interaction by objectively recording social events happening between individuals at one point in time via different communication channels such as calls, texts, and proximity recordings by Bluetooth scans, and hereby also record interactions that individuals do not remember or perceive as being important [20]. Self-reports of social relations often reflect individual evaluations of social ties that exist over a longer time span [21]. Nevertheless, despite the distinct differences inherent in the two types of data, it is possible that there is an overlap as both methods reflect social connectivity, and using both might be complementary for measuring individual level social relationships [20, 22]. In order to use smartphones as measurement tools to study individual social relations, we propose to derive relevant smartphone measures of social relations based on theoretical concepts, and to further explore the content of such smartphone measures.

## Social integration

Social integration has been widely studied in relation to well-being and health outcomes, and the literature is rich in various ways of defining and measuring this construct [23–27]. This multifaceted construct has been conceptualized as the extent to which individuals are connected within and participate in a broad range of social relations and activities reflecting how well the individual is connected within a social network [25–27]. Two concepts can be used to operationalize social integration: the *social network size* and the *social role diversity* [26]. The Social network size is defined as the number of social ties connected to an individual, and hence reflect how well an individual is connected within a social network. In contrast to counting the sheer number of social ties, social role diversity indicates the number of different social roles surrounding the individual. A person might engage in social interaction with a friend in one social context, a sibling in a second, and with a parent in yet another context. Fig 1 depicts social network size and the social role diversity for one individual (ego). The figure shows a network size of seven, but a social role diversity of five as some of the social ties are characterized by having the same social role, e.g. friendship. Nonetheless, both measures reflect the extent to which an individual participates in different contexts involving various social activities, and hereby also provide insight into information about one’s level of social integration. Existing survey instruments evaluating diversity include social roles relating to both family, non-family roles such as friends and neighbours and roles from local community activities [1, 27, 28]. Given that having a high social role diversity relates to being in contact with a wide range of different people [26], a high social role diversity might also be reflected in social interaction behavior via smartphones, with social integration in this context meaning a larger communication network. Nevertheless, this hypothesis is largely empirically untested. We identified one study investigating network size of call patterns, but this smartphone measure was not directly compared against a self-reported measure [29].



**Fig 1. Illustration of social network size, social role diversity, social ties (tie strength, and reciprocity).** Fig 1 illustrates definitions of social relation concepts. *The Diversity* is defined as the number of different social roles (e.g. friend, brother, mother), *The Network Size* is defined as the number of social ties (all arrows in the figure), *The Tie Strength* is defined as a continuum of closeness in a relationship ranging from weak ties to strong ties (Thickness of arrow). *The Reciprocity* is defined as the mutuality in a social tie and describes a social tie between two individuals where the tie is directed both ways (bi-directional arrow).

<https://doi.org/10.1371/journal.pone.0200678.g001>

### Tie strength

Social ties can be characterized by their strength. Tie strength is a continuum of closeness in a relationship, ranging from weak ties (narrow arrows in Fig 1), to strong ties which are ties to people with whom the individual has an intimate relationship with, for example, a close friend or a family member (bold arrows in Fig 1). Tie strength has been defined as “.. a combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and reciprocal services that characterizes the tie” [30], and hence used indicators of tie strength include contact frequency, tie duration, and tie reciprocity [31]. The *Contact frequency* is the sum of social interactions in a defined time period, which can be carried out in a face-to-face encounter or in a mediated encounter, e.g. social contact via a smartphone. The *Duration* considers the duration of social interactions, where longer duration is an indicator of strong ties [30, 32]. *Tie reciprocity* describes the mutuality in a social tie (illustrated by the bi-directional arrow in Fig 1), where mutual relationships are indicators of strong ties [30]. These tie strength indicators are considered closely related as they measure the same underlying concept of tie strength [33]. The indicators can be described as interactional network characteristics as they reflect the extent of interactions a respondent has with alters, i.e. surrounding individuals. As such these indicators describe the structure in which respondent’s perception of closeness in a social tie is likely to occur. In the relatively sparse literature comparing smartphone data to survey data, predicting self-reported closeness in a social tie with smartphone data has most often been done [20, 22, 29, 34]. Studies using measures such as frequency of calls [34], reciprocity in call patterns [35], as well as face-to-face proximity recorded with Bluetooth data [20, 22] have been able to predict self-reported closeness in a social tie. As smartphone data represent count data, they might be superior in evaluating sheer structural aspects of tie strength for example contact frequency. In the paper, we will emphasize the structural aspects of tie strength as we consider smartphone data adequate for measuring such aspects.

In this paper, we conduct conceptual work aiming to understand aspects of social integration, and tie strength captured in smartphone data. We hypothesize that derived information on social interactions collected continuously via smartphones during a one-month period is associated with self-reported measures of social integration and tie strength. We do not attempt to validate smartphone measures against a golden survey standard of social relations measurement as such does not exist.

## Material and methods

### Study population

We used data from the Copenhagen Network Study, which was established to study social activity and behaviors based on smartphone data among young adults [36]. In September 2013, 3,329 undergraduate students at a large technical university in Denmark were invited to participate in the study of which 979 students accepted the invitation (response rate = 29%). The majority of the participating students were freshmen students (60%). All participants received a smartphone (LG NEXUS 4) in which they inserted their private SIM-card to make it their primary communication device. The smartphone was running customized software recording all outgoing and ingoing call and text messages with related timestamps as well as unique identifiers for each contacted alter. Before receiving the smartphone, the students completed a baseline questionnaire on self-reported social relations. A detailed description of the high-resolution smartphone data collection can be seen in Stopczynski et al. 2014 [36]. The participants were recruited continuously throughout the year, and the used smartphone data were recorded continuously in one month from receiving the smartphone and responding to the baseline questionnaire. We excluded individuals with no information on self-reported social relations (N = 33), and with missing smartphone recordings (N = 209) yielding a total sample of 737 participants. Missing smartphone data was not related to gender (chi-squared test, p-value = 0.80), but was related to age where younger students were more likely to have their smartphone data recorded (t test, p-value = 0.026).

### Measurements

To measure the concepts of social integration and tie strength, we used validated survey items from the Copenhagen Social Relations Questionnaire [37], and derived smartphone data from both call and text messages. Table 1 shows a summary of the measures used.

**Social integration.** As a measure of the *Social role diversity*, we assessed the number of six different social roles for which the participant self-reported frequent face-to-face contact using the item: How often are you together with any of the following people (mother, father, siblings,

**Table 1. Overview of concepts, operationalization, data types and measures.**

Concept	Operationalization	Data type	Measure
Social integration	Social role diversity	Survey data	Number of social roles with frequent contact (contact one to three times per month or more) both face-to-face and non-face-to-face with roles; Mother, Father, Friends, Partner, Siblings, Extended family)
	Social network size	Smartphone data	Number of alters called and texted at least once during a month
Tie strength	Total contact frequency	Survey data	Total face-to-face and non-face-to-face contact frequency with all roles (Mother, Father, Friends, Partner, Siblings, Extended family)
	Contact frequency, Tie duration, Tie reciprocity	Smartphone data	Frequency of calls and texts per alter per month, Total duration of calls per month, number of alters with reciprocated activity (both placing and receiving at least one call or text from the same alter).

<https://doi.org/10.1371/journal.pone.0200678.t001>

extended family, partner, and friends), who you do not live with? (Response code: Several days a week; About once a week; One to three times a month; Less often than once a month; Never; Have no; Live with). The family and non-family social roles were considered relevant for a population of college students who are in a transitional stage of life increasingly creating close relations with peers, but still relying on close family members for support, mutual confiding and information [38]. A contact frequency of one to three times a month or more as well as co-habitation was considered frequent contact corresponding to active social role participation defined elsewhere [26]. Having less frequent social contact or having no social role was used as an indicator that no role was present. The derived variable of the number of social roles was categorized in intervals of two. One participant had no social roles and was grouped in the lowest category. Similarly, we also assessed the social role diversity in non-face-to-face contacts for the same six social roles using the item: How often do you have contact with any of the following people without seeing them (e.g. by telephone, letters, e-mail, SMS)? From the smartphone data, we derived a measure of the *Social network size* by counting the number of different alters that the participant had interacted with at least once in a month via either placed or received calls and text messages. The social network size values were grouped in intervals of ten.

**Tie strength.** From the self-reported face-to-face and non-face-to-face contact frequency survey items described above, we derived a composite measure indicating the *Total contact frequency* with all social roles. We summed the six contact frequency items containing five categories (coded 0–4) on a scale from 0–24 where a score of 24 indicated reporting “Several days a week” or “Live with” for all social roles and a score of 0 indicated reporting “Never” or “Have no” for all roles. In order to maintain important information on strong social ties, individuals reporting “live with” was grouped in the highest contact frequency category. The summary scale was grouped in five-interval categories. We developed a similar measure for the total non-face-to-face contact frequency. From the smartphone data, we also derived a measure of total contact frequency by summing of the total number of calls and text messages during a month normalized by the number of alters in the social network. The call interaction variable was grouped in intervals of three, and the text interaction variable was grouped in intervals of fifteen. Further, to construct an indicator of the *total duration* of social interactions, we summed the duration of calls for each individual during a month and grouped this variable in one-hour intervals. Non-received calls were excluded. As a measure of *tie reciprocity*, we counted for each respondent the number of ties with reciprocated smartphone activity defined as having both placed and received at least one call or text message from the same alter. These variables were grouped in three intervals for call reciprocity and six intervals for text message reciprocity. Nine participants did not have call activity and hence were excluded from this variable as reciprocation in this situation was not meaningful. We did not have survey data available indicating duration and reciprocity in social ties.

### Analytical strategy

We investigated distributions of age, gender, self-reported and derived smartphone measures of social integration and tie strength in the study population. To assess the association between the self-reported and smartphone variables of social integration, we estimated odds ratios and 95% confidence intervals using logistic regression models with the highest social role diversity category as outcome category (5–6 social roles). Likewise, associations between smartphone measures of contact frequency, duration and reciprocity and self-reported measures of total contact frequency were evaluated using the highest self-reported total contact frequency as outcome, i.e. scoring between 20–24 on the derived total contact frequency summary scale.

The evidence of trends in associations, e.g. whether the odds of having high social role diversity was increasing with a larger call network, was assessed by including the categorized smartphone measure as continuous in the logistic regression model. As social relations and smartphone usage vary with age and gender [39, 40], and as calling behavior of young adults appears to be influenced by whether they live with their parents [38], all models were adjusted for age, gender and co-habitation (living with at least one of the six social roles). We conducted the following sensitivity analyses: 1) Counting co-habitation as an active social role might categorize some individuals as having a high social integration despite having few friends, e.g. individuals living at home with family members but indicating infrequent contact with friends. Hence, we conducted a sensitivity analysis excluding co-habiting individuals to assess whether the results were robust to this categorization. 2) To assess whether the results were robust to dichotomizing self-reported social relation variables in high versus low for use in logistic regression models, we also conducted linear regression analyses using a continuous version of the measures. 3) As the smartphone measure of social network size was in risk of being wrongly categorized due to service call and alike, we restricted the measure to a definition where network size was counted as the number of unique persons called/texted at least three times in a sensitivity analysis. All analyses were conducted using statistical software R version 3.3.3. [41].

### Ethics statement

All participants gave informed consent and were able to withdraw from the study at any time. All data were used anonymously, and The Copenhagen Network Study was approved by the Danish Data Protection Agency (approval number: 2012-41-0664).

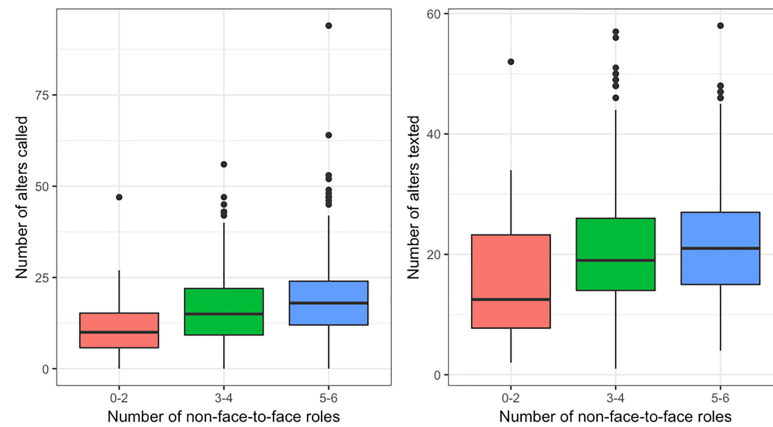
## Results

### Descriptive statistics

The majority of the population were males (77%), and the mean age was 21.6 year (SD: 2.6) ranging from 18 to 46 years. Distributions of smartphone measures can be seen in [S1 Table](#). The majority of participants were in call (38.8%) and text (40.4%) contact at least once with between 11 and 20 different alters in a month. Further, the majority of the study population had between 4–6 call interactions (51.6%) and 0–14 text interactions (40.4%) per alter in a month. The majority of the study population had a call duration of 3 hours or more during a month (38.9%), and between 4–6 reciprocated call (33.7%) and more than 18 reciprocated text message ties (32.7%). Women had larger text message networks, a higher text message frequency, longer call duration and a higher text message reciprocity compared to men. The mean age was higher among participants with a large call network and a long call duration, but lower for participants with a high text message frequency. Distributions of self-reported survey measures can be seen in [S2 Table](#). 41.7% and 56.9% of the study population had a high social role diversity in face-to-face contact and non-face-to-face contact, respectively. 6.1% and 15.5% scored between 20–24 on the total face-to-face and non-face-to-face contact frequency summary scale, respectively. 315 individuals reported to be co-habiting with a social role where the majority of these (46%) were living with their parents. Participants with a higher social role diversity and self-reported contact frequency appeared to be younger than participants with lower contact frequency and role diversity.

### Social integration

[Fig 2](#) shows medians of smartphone network size in groups of self-reported social role diversity. From the figure it is apparent, as hypothesized, that the median of alters contacted via



**Fig 2. Boxplots of the associations between self-reported social role diversity and smartphone derived measures of social network size.** Fig 2 illustrates medians (midline of box), 25<sup>th</sup> and 75<sup>th</sup> percentiles (upper and lower edge of box) of smartphone measured social network size in groups of self-reported non-face-to-face social role diversity.

<https://doi.org/10.1371/journal.pone.0200678.g002>

smartphones is increasing with the number of self-reported social roles. This tendency is confirmed when adjusting for age, gender and co-habitation (Table 2). Participants having contact with more than 30 different alters via calls and text messages have more than two times higher odds (Call: OR = 2.58, 95%CI:1.50;4.44) (Text: OR = 2.55, 95%CI: 1.44;4.50) of also having frequent non-face-to-face contact with 5–6 social roles compared to participants who have call and text message contact with only 0–10 alters. The same pattern appears for self-reported face-to-face diversity (Table 2). There is strong evidence for trends in the reported associations suggesting that the odds ratios for having a high social role diversity is increasing with a larger call and text message network. Using a continuous version of the self-reported social relation variable did not change the overall conclusion of these findings (S3 Table). Further, excluding co-habiting individuals as well as using a more restricted definition of call and text message network size did not change the conclusion of the results although the association appeared to be more pronounced for non-face-to-face diversity (S4 and S5 Tables).

**Table 2. Associations between smartphone and self-reported measures of social integration in a population of 737 young adults.**

Smartphone social network size per month	Total population	High social role diversity (Frequent face-to-face contact with 5–6 social roles)		High social role diversity (Frequent non-face-to-face contact with 5–6 social roles)	
	N (%)	OR	95%CI	OR	95%CI
<i>Number of alters called</i>					
0–10 alters	186 (25.2)	1	(Ref)	1	(Ref)
11–20 alters	286 (38.8)	1.26	(0.85;1.87)	1.91	(1.30;2.79)
21–30 alters	181 (24.6)	1.50	(0.97;2.32)	2.78	(1.80;4.28)
More than 30 alters	84 (11.4)	1.82	(1.05;3.14)	2.58	(1.50;4.44)
P-value (test for trend)		0.018		<0.0001	
<i>Number of alters texted</i>					
0–10 alters	92 (12.5)	1	[Ref]	1	(Ref)
11–20 alters	298 (40.4)	1.99	(1.18;3.34)	1.64	(1.02;2.64)
21–30 alters	223 (30.3)	2.26	(1.31;3.88)	2.20	(1.33;3.63)
More than 30 alters	124 (16.8)	2.31	(1.26;4.22)	2.55	(1.44;4.50)
P-value (test for trend)		0.014		0.0005	

OR = Odds ratio, 95%CI = 95% confidence interval. All OR adjusted for age, gender and co-habitation.

<https://doi.org/10.1371/journal.pone.0200678.t002>



### Tie strength

A higher smartphone contact frequency of calls and text messages is also strongly associated with self-reported face-to-face contact frequency. Having more than 9 calls per alter is associated with six times higher odds (OR = 6.53, 95%CI:1.98;21.47) of also scoring high on the self-reported contact frequency scale (20–24) compared to having between 0–3 calls per alter. The same tendency appeared for the self-reported non-face-to-face contact frequency, which also showed strong associations with call interaction frequency. Longer call duration and higher reciprocity in call and text message communication was associated with reporting a high face-to-face contact frequency, and longer call duration and higher reciprocity in call communication was associated with reporting a high non-face-to-face contact frequency (Table 3). Using

**Table 3. Associations between smartphone and self-reported measures of tie strength in a population of 737 young adults.**

Smartphone measures per month	Total population	High face-to-face contact frequency (summary score = 20–24)		High non-face-to-face contact frequency (summary score = 20–24)	
	N (%)	OR	95%CI	OR	95%CI
<i>Frequency of call interactions per alter</i>					
0–3 calls	221 (30.0)	1	(Ref)	1	(Ref)
4–6 calls	380 (51.6)	2.85	(1.13;7.16)	2.36	(1.34;4.16)
7–9 calls	96 (13.0)	2.51	(0.77;8.22)	3.10	(1.54;6.28)
More than 9 calls	40 (5.4)	6.53	(1.98;21.47)	5.82	(2.55;13.28)
P-value (test for trend)		0.004		<0.0001	
<i>Frequency of text interactions per alter</i>					
0–14 texts	298 (40.4)	1	(Ref)	1	(Ref)
15–29 texts	240 (32.6)	5.92	(1.94;18.10)	1.80	(1.09;2.98)
30–44 texts	100 (13.6)	5.48	(1.57;19.15)	2.01	(1.08;3.74)
More than 45 texts	99 (13.4)	11.44	(3.63;36.04)	1.58	(0.83;3.02)
P-value (test for trend)		<0.0001		0.072	
<i>Call duration</i>					
up to 1 hr	160 (21.7)	1	(Ref)	1	(Ref)
1–2 hrs	150 (20.4)	2.43	(0.89;6.66)	1.86	(0.92;3.75)
2–3 hrs	140 (19.0)	2.86	(0.98;8.34)	1.86	(0.90;3.85)
More than 3 hrs	287 (38.9)	3.02	(1.17;7.77)	2.84	(1.51;5.35)
P-value (test for trend)		0.029		0.001	
<i>Call reciprocity*</i>					
0–3 reciprocated ties	145 (19.9)	1	(Ref)	1	(Ref)
4–6 reciprocated ties	245 (33.7)	1.31	(0.50;3.39)	1.44	(0.72;2.88)
7–9 reciprocated ties	167 (22.9)	2.60	(0.98;6.90)	2.61	(1.30;5.26)
More than 9	171 (23.5)	2.04	(0.73;5.74)	3.47	(1.74;6.93)
P-value (test for trend)		0.069		<0.0001	
<i>Text reciprocity</i>					
0–6 reciprocated ties	82 (11.1)	1	(Ref)	1	(Ref)
7–12 reciprocated ties	205 (27.8)	2.23	(0.61;8.12)	1.72	(0.75;3.92)
13–18 reciprocated ties	209 (28.4)	1.89	(0.50;7.13)	1.56	(0.67;3.59)
More than 18	241 (32.7)	4.09	(1.09;15.31)	2.23	(0.98;5.08)
P-value (test for trend)		0.040		0.079	

OR = Odds ratio, 95%CI = 95% confidence interval. All OR adjusted for age, gender, and co-habitation.

\*9 observations excluded because of no calling activity.

<https://doi.org/10.1371/journal.pone.0200678.t003>

a continuous version of the measures did not change these conclusions although there was no association between call duration and face-to-face contact frequency (S3 Table).

## Discussion

In this study following more than 700 young adults with survey data and continuous smartphone recordings, we attempted to investigate whether smartphone measures of interactions would be associated with specific survey measures based on the assumption that they to some extent would measure the same underlying theoretical concepts. Confirming this hypothesis, we found considerable associations between self-reports and smartphone measures of social relations. Being in contact with a high number of different individuals via both calls and text messages was associated with reporting to have frequent contact with a high number of social roles. Further, having a high total frequency of calls and text messages, high duration of calls, and high call and text reciprocity was also associated with self-reporting a high contact frequency.

## Social integration

To the best of our knowledge, this is the first study to investigate associations between self-reported social role diversity and social network size in phone communication. Miritello et al. (2013) investigated social network size in individual networks among 20 million mobile users [29]. They did not compare this measure against self-reports but detected similar social connectivity patterns often found in self-reported social networks, where individuals have an upper limit for the number of alters included in the network. One study investigating diversity in communication behavior found that poor mental health was related to smaller communication networks [8].

Even though the current study shows promising results for measuring social integration aspects with smartphone data, one should be aware that the used self-report measure did not evaluate all aspects of social integration. Contrary to other survey instruments evaluating social role diversity [27], the used survey instruments did not measure roles relating to leisure or local community activities, and hence we were not able to consider this aspect of social integration. One study found associations between self-reported phone use and spending leisure time with others with the underlying hypothesis that individuals keep in contact via mobile phones before meeting and organizing events [42]. However, further research is needed to clarify whether objectively measured smartphone communication is associated with leisure and local community activities. Other dimensions of social integration which have rarely been investigated for a larger population due to data limitations of survey self-reports are concepts such as density in social networks, and centrality in social networks [43]. These measures are difficult to obtain on a large scale with self-reports as this requires extensive mapping of whole (i.e. socio-centric) networks. Smartphone data provide possibilities to also investigate social networks beyond the individual level if data is collected with a considerable high coverage within a somewhat specifically defined environment, e.g. neighborhood, workplace, or educational institution. In the study we were not able to consider such measures because of the relatively low response rate.

The differences in the two measures used to operationalize social integration, i.e. social role diversity and social network size should also be noted. The survey items used were only able to evaluate contact with up to six different social roles, whereas there might be considerable more diversity in a network of communication. Further, the number of alters included in the social role “friend” vary from person to person meaning that the diversity in phone contacts might differ more than what can be captured with the survey item of social role diversity used. This

point might also be especially relevant considering the population under study consisting of college students who possibly have a high contact frequency with a range of friends. Social roles can be viewed as sets of behavioral contingencies which are a result of the social interactions in one's proximal environment [27]. Because the social environment of young adulthood in college is likely primarily comprised of a milieu of peers, the behavioural expectations and cues that emerge from learned behaviours in this social environment are largely based on expectancies formed from peer-to-peer interactions. As such, it is likely that college students maintain fewer social roles in comparison to adults during this transitional developmental time period of college as they spend most of their time interacting socially with peers rather than engaging in many diverse social roles. Hence, it is possible that smartphone measures are actually a more valid representation of the construct of social integration for a population of young adults given that they are in college and they interact with peers via smartphone devices. Challenges using the smartphone for estimating social network size, nevertheless, also include a risk of overestimation due to service calls and alike. We tried to minimize this measurement error by conducting sensitivity analyses restricting social contacts called at least three times.

### Tie strength

We also found considerable associations between self-reported and smartphone recorded contact frequency. Contact frequency via smartphones has been found in other studies to predict tie strength. In a study by Saramäki et al., they found that the frequency of calls to a specific alter predict self-reported closeness as well as self-reported number of days since last face-to-face encounter with the same individual [34]. Another study found that self-reported emotional closeness was related to having short time between phone communication events [44]. Hence, there appears to build-up evidence that contact frequency via phone communication may be a good predictor of tie strength. Predicting self-reports of tie strength with reciprocity in and duration of smartphone communication has rarely been done, but we found convincing evidence for an association between the two data sources with respect to both reciprocity and duration. In the study by Miritello et al. distributions of tie strength as duration of phone calls was investigated [29]. They did not compare this measure against self-reports, but they detected constraints on the number of strong ties in large networks suggesting similar social connectivity patterns often found in self-reported social networks, where individuals have an upper limit for strong ties. Although frequency, duration and reciprocity in call and text patterns might be good proxies for functional dimensions of tie strength such as closeness and intimacy these functional dimensions might be complex to study directly with smartphone data as evaluation of such aspects would require access to the content of the calls and text messages.

### Privacy and data access

Along with the increasing usage of smartphone data, privacy issues is a concern especially when merged with sensitive data concerning health and well-being [45]. On the other hand, the scientific community may also have obligations to take on and explore new possibilities in data and technology that can potentially contribute to the improvement of the public health of populations [46, 47]. Whereas collection of smartphone data requires very little efforts of the study participants, it still remains a challenge to access the digital traces from smartphones. Most national phone companies hold large scale smartphone generated data that could constitute basis for research. Nevertheless, companies can be reluctant to share these data due to privacy issues or own commercial interests, and often the research community does not have direct access to large scale smartphone data [48]. Digital traces from social media also

constitute valuable information on social relations. Nevertheless, one should keep in mind that the representativeness of social media data might be compromised by user preferences for the specific social media [49, 50], whereas this might be less of a problem with smartphone data such as calls and text messages as the coverage in most of the western world is high [40]. Collecting smartphone data a priori with informed consent as done in the present study allows direct access to relevant smartphone interactions as well as merging with other survey and health administration data, but this approach is feasible on a smaller scale.

### Strengths and limitations

This study used data from one of the largest databases containing detailed information on both continuous smartphone recordings around the clock linked to relevant self-reports from surveys. Nevertheless, the response rate of 29% should be taken into consideration when interpreting the results. Unfortunately, we did not have data available to further explore characteristics of non-responders. Further, one should be aware that the study population does not constitute a random sample of the population but reflect a selected population of young adults attending higher education, who were likely to interact socially. The pattern of and motivations for smartphone use differ by age [39], and hence the results might be less generalizable to older age groups.

### Conclusion

The use of digital traces from social interactions in the health and social science research is inevitable. We have shown a considerable overlap between self-reported and smartphone derived measures of social relations and conclude that smartphone data hold promising potential for a detailed measurement of social interaction and social relations, which can be used as supplementary information to established survey measures.

### Supporting information

**S1 Table. Associations between age, gender and smartphone measures of social relations in a population of 737 young adults.**

(DOCX)

**S2 Table. Associations between age, gender and self-reported measures of social relations in a population of 737 young adults.**

(DOCX)

**S3 Table. Linear regression of associations between continuous smartphone measures and self-reported social relation variables.**

(DOCX)

**S4 Table. Associations between smartphone and self-reported measures of social integration in a population of 737 young adults excluding co-habiting individuals.**

(DOCX)

**S5 Table. Associations between self-reported and smartphone measures of social integration in a population of 737 young adults. Restricted social network size to a minimum of three social interactions per unique alter.**

(DOCX)

## Acknowledgments

We thank The University of Copenhagen 2016-initiative and SensibleDTU.

## Author Contributions

**Conceptualization:** Agnete S. Dissing, Cynthia M. Lakon, Naja H. Rod, Rikke Lund.

**Data curation:** Agnete S. Dissing, Thomas A. Gerds.

**Formal analysis:** Agnete S. Dissing.

**Funding acquisition:** Naja H. Rod, Rikke Lund.

**Investigation:** Agnete S. Dissing, Cynthia M. Lakon, Rikke Lund.

**Methodology:** Agnete S. Dissing, Cynthia M. Lakon, Thomas A. Gerds, Naja H. Rod, Rikke Lund.

**Project administration:** Agnete S. Dissing, Rikke Lund.

**Software:** Agnete S. Dissing, Thomas A. Gerds.

**Supervision:** Cynthia M. Lakon, Thomas A. Gerds, Naja H. Rod, Rikke Lund.

**Writing – original draft:** Agnete S. Dissing.

**Writing – review & editing:** Agnete S. Dissing, Cynthia M. Lakon, Thomas A. Gerds, Naja H. Rod, Rikke Lund.

## References

1. Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. *Am J Epidemiol.* 1979; 109(2):186–204. PMID: [425958](#)
2. Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: a meta-analytic review. *PLoS Med* 2010; 7(7):e1000316. <https://doi.org/10.1371/journal.pmed.1000316> PMID: [20668659](#)
3. House JS, Landis KR, Umberson D. Social relationships and health. *Science.* 1988; 241(4865):540–5. PMID: [3399889](#)
4. Cohen S. Social relationships and health. *Am Psychol.* 2004; 59(8):676–84. <https://doi.org/10.1037/0003-066X.59.8.676> PMID: [15554821](#)
5. Smith KP, Christakis NA. Social networks and health. *Annu Rev Sociol.* 2008; 34(1):405–29.
6. Ayers JW, Althouse BM, Dredze M. Could behavioral medicine lead the web data revolution? *JAMA.* 2014; 311(14):1399–400. <https://doi.org/10.1001/jama.2014.1505> PMID: [24577162](#)
7. Lazer D, Pentland AS, Adamic L, Aral S, Barabasi AL, Brewer D, et al. Life in the network: the coming age of computational social science. *Science.* 2009; 323(5915):721–3. <https://doi.org/10.1126/science.1167742> PMID: [19197046](#)
8. Madan A, Cebrian M, Lazer D, Pentland A. Social sensing for epidemiological behavior change. *ACM press.* 2010 [cited New York (NY):291.
9. Oliver N, Matic A, Frias-Martinez E. Mobile network data for public health: opportunities and challenges. *Front Public Health* 2015; 3(189).
10. Onnela J-P, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, et al. Structure and tie strengths in mobile communication networks. *PNAS.* 2007; 104(18):7332–6. <https://doi.org/10.1073/pnas.0610245104> PMID: [17456605](#)
11. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al., editors. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*; 2014 13–17 September; Seattle, Washington, USA: ACM.
12. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-Generation Psychiatric Assessment: Using Smartphone Sensors to Monitor Behavior and Mental Health. *Psychiatr Rehabil J.* 2015; 38(3):218–26. <https://doi.org/10.1037/prj0000130> PMID: [25844912](#)

13. Lane ND, Miluzzo E, Lu H, Peebles D, Choudhury T, Campbell AT. A survey of mobile phone sensing. *IEEE Commun Mag*. 2010; 48(9):140–50.
14. Torous J, Kiang MV, Lorme J, Onnela J-P. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. *JMIR mental health*. 2015; 3(2):e16–e.
15. Onnela J-P, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology*. 2016; 41(7):1691–6. <https://doi.org/10.1038/npp.2016.7> PMID: [26818126](https://pubmed.ncbi.nlm.nih.gov/26818126/)
16. Sun L, Axhausen KW, Lee D-H, Huang X. Understanding metropolitan patterns of daily encounters. *PNAS*. 2013; 110(34):13774–9. <https://doi.org/10.1073/pnas.1306440110> PMID: [23918373](https://pubmed.ncbi.nlm.nih.gov/23918373/)
17. Madan A, Farrahi K, Gatica-Perez D, Pentland AS. Pervasive sensing to model political opinions in face-to-face networks. In: Lyons K, Hightower J, Huang EM, editors. *Pervasive Computing Pervasive 2011 Lecture Notes in Computer Science*. 6696. Berlin, Heidelberg: Springer; 2011.
18. Eagle N, Macy M, Claxton R. Network diversity and economic development. *Science*. 2010; 328(5981):1029–31. <https://doi.org/10.1126/science.1186605> PMID: [20489022](https://pubmed.ncbi.nlm.nih.gov/20489022/)
19. Karsai M, Perra N, Vespignani A. The emergence and role of strong ties in time-varying communication networks. *arXiv preprint arXiv:13035966*. 2013.
20. Eagle N, Pentland AS, Lazer D. Inferring friendship network structure by using mobile phone data. *PNAS*. 2009; 106(36):15274–8. <https://doi.org/10.1073/pnas.0900282106> PMID: [19706491](https://pubmed.ncbi.nlm.nih.gov/19706491/)
21. Wd Nooy. Structure from interaction events. *BD&S*. 2015; 2(2):1–4.
22. Zhao H, Zhou H, Yuan C, Huang Y, Chen J. Social Discovery: Exploring the Correlation Among Three-Dimensional Social Relationships. *IEEE Transactions on Computational Social Systems*. 2016; 2(3):77–87.
23. Berkman LF, Glass T. Social integration, social networks, social support, and health. In: Berkman LF, Kawachi I, editors. *Social epidemiology*. 12000. p. 137–73.
24. Berkman LF, Glass T, Brissette I, Seeman TE. From social integration to health: Durkheim in the new millennium. *Soc Sci Med*. 2000; 51(6):843–57. PMID: [10972429](https://pubmed.ncbi.nlm.nih.gov/10972429/)
25. House JS, Umberson D, Landis KR. Structures and processes of social support. *Annu Rev Sociol*. 1988; 14:293–318.
26. Brissette I, Cohen S, Seeman TE. Measuring social integration and social networks. In: Cohen S, Underwood LG, Gottlieb BH, editors. *Social support measurement and intervention: A guide for health and social scientists*. New York: Oxford University Press; 2000.
27. Cohen S, Brissette I, Skoner DP, Doyle WJ. Social integration and health: The case of the common cold. *J Soc Struc*. 2000; 1(3):1–7.
28. Orth-Gomer K, Johnson JV. Social network interaction and mortality: a six year follow-up study of a random sample of the Swedish population. *J Chronic Dis*. 1987; 40(10):949–57. PMID: [3611293](https://pubmed.ncbi.nlm.nih.gov/3611293/)
29. Miritello G, Moro E, Lara R, Martínez-López R, Belchamber J, Roberts SG, et al. Time as a limited resource: Communication strategy in mobile phone networks. *Social Netwks*. 2013; 35(1):89–95.
30. Granovetter MS. The strength of weak ties. *Amer J Sociol*. 1973; 78(6):1360–80.
31. Marsden PV, Campbell KE. Measuring tie strength. *Soc Forces*. 1984; 63(2):482–501.
32. Hall JA. How many hours does it take to make a friend? [published online ahead of print March 15, 2018]. *J Soc Pers Relat*.
33. Hill RA, Dunbar RI. Social network size in humans. *Hum Nat*. 2003; 14(1):53–72. <https://doi.org/10.1007/s12110-003-1016-y> PMID: [26189988](https://pubmed.ncbi.nlm.nih.gov/26189988/)
34. Saramäki J, Leicht E, López E, Roberts SG, Reed-Tsochas F, Dunbar RI. Persistence of social signatures in human communication. *PNAS*. 2014; 111(3):942–7. <https://doi.org/10.1073/pnas.1308540110> PMID: [24395777](https://pubmed.ncbi.nlm.nih.gov/24395777/)
35. Zhang H, Dantu R. Predicting social ties in mobile phone networks. *IEEE International Conference on Intelligence and Security Informatics2010*. p. 25–30.
36. Stopczynski A, Sekara V, Sapiezynski P, Cuttone A, Madsen MM, Larsen JE, et al. Measuring large-scale social networks with high resolution. *PLoS One*. 2014; 9(4):e95978. <https://doi.org/10.1371/journal.pone.0095978> PMID: [24770359](https://pubmed.ncbi.nlm.nih.gov/24770359/)
37. Lund R, Nielsen LS, Henriksen PW, Schmidt L, Avlund K, Christensen U. Content validity and reliability of the Copenhagen Social Relations Questionnaire. *J Aging Health*. 2014; 26(1):128–50. <https://doi.org/10.1177/0898264313510033> PMID: [24584264](https://pubmed.ncbi.nlm.nih.gov/24584264/)
38. Chen Y-F, Katz JE. Extending family to school life: College students' use of the mobile phone. *Int J Hum Comput Stud*. 2009; 67(2):179–91.

39. Forgays DK, Hyman I, Schreiber J. Texting everywhere for everything: Gender and age differences in cell phone etiquette and use. *Comput Human Behav.* 2014; 31:314–21.
40. Pew Research Center. Mobile Fact Sheet 2018 [Available from: <http://www.pewinternet.org/fact-sheet/mobile/>].
41. R Core Team. R: A Language and Environment for Statistical Computing [computer program]. R Foundation for Statistical Computing Vienna, Austria, 2016–2017.
42. Campbell SW, Kwak N. Mobile communication and social capital: An analysis of geographically differentiated usage patterns. *New Media Soc.* 2010; 12(3):435–51.
43. Uchino BN. *Social support and physical health: Understanding the health consequences of relationships.* New Haven CT: Yale University Press; 2004.
44. Roberts SG, Dunbar RI. Communication in social networks: Effects of kinship, network size, and emotional closeness. *Pers Relatsh.* 2011; 18(3):439–52.
45. Vayena E, Salathé M, Madoff LC, Brownstein JS, Bourne PE. Ethical challenges of big data in public health. *PLoS Comput Biol.* 2015; 11(2):e1003904. <https://doi.org/10.1371/journal.pcbi.1003904> PMID: 25664461
46. Gange SJ, Golub ET. From Smallpox to Big Data: The Next 100 Years of Epidemiologic Methods. *Am J Epidemiol.* 2015; 183(5):423–6. <https://doi.org/10.1093/aje/kwv150> PMID: 26443419
47. Mooney SJ, Westreich DJ, El-Sayed AM. Epidemiology in the Era of Big Data. *Epidemiology.* 2015; 26(3):390–94. <https://doi.org/10.1097/EDE.0000000000000274> PMID: 25756221
48. Manovich L. Trending: The promises and the challenges of big social data. In: Gold MK, Minneapolis MN, editors. *Debates in the digital humanities: The University of Minnesota Press.* [Online] 2011.
49. Ruths D, Pfeffer J. Social media for large studies of behavior. *Science.* 2014; 346(6213):1063–4. <https://doi.org/10.1126/science.1256213> PMID: 25430759
50. Schober MF, Pasek J, Guggenheim L, Lampe C, Conrad FG. Social media analyses for social measurement. *Public Opin Q.* 2016; 80(1):180–211. <https://doi.org/10.1093/poq/nfv048> PMID: 27257310