

Weiler, Michael; Jansen, Nora; Hinz, Oliver

Article

Can we measure the structural dimension of social capital with digital footprint data? An assessment of the convergent validity of an indicator extracted from digital footprint data

Schmalenbach Journal of Business Research (SBUR)

Provided in Cooperation with:

Schmalenbach-Gesellschaft für Betriebswirtschaft e.V.

Suggested Citation: Weiler, Michael; Jansen, Nora; Hinz, Oliver (2024) : Can we measure the structural dimension of social capital with digital footprint data? An assessment of the convergent validity of an indicator extracted from digital footprint data, Schmalenbach Journal of Business Research (SBUR), ISSN 2366-6153, Springer, Heidelberg, Vol. 76, Iss. 2, pp. 159-195, <https://doi.org/10.1007/s41471-024-00180-8>

This Version is available at:

<https://hdl.handle.net/10419/298797>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Can We Measure the Structural Dimension of Social Capital with Digital Footprint Data?—An Assessment of the Convergent Validity of an Indicator Extracted from Digital Footprint Data

Michael Weiler · Nora Jansen · Oliver Hinz 

Received: 25 August 2021 / Accepted: 30 January 2024 / Published online: 4 March 2024
© The Author(s) 2024

Abstract Network analytical metrics often seek to capture the structural dimension of social capital, but such data collections using traditional social research tools often suffer from biases like interviewer effects and are usually only suitable to study small groups of participants. Digital sources of social relations might offer great potential for facilitating such measures though, because they digitally store unprecedented amounts of relational data, free from the limitations associated with self-reported data. This study therefore compares individual node degrees collected through a contact diary (i.e., overall-social capital), and a counterpart extracted from digital footprint data from the social media network, Facebook (i.e., social media network-social capital). The findings suggest that researchers conducting empirical studies involving the concept thus should not ignore social media network-social capital as a practical alternative measure of overall-social capital; it provides a sound approximation but only after controlling for other influential factors. In particular, our results highlight that the usability of the digital social capital metric is conditional on the three-way interaction between the variables gender, age, and social media network-social capital. Thus, the evidence from our study, in turn, also intimates that individuals act heterogeneously in the digital sphere with respect to their networking behaviour.

Michael Weiler · Nora Jansen · ✉ Oliver Hinz
Faculty of Economics and Business Administration, Goethe University Frankfurt,
Theodor-W.-Adorno-Platz 4, 60323 Frankfurt am Main, Germany
E-Mail: ohinz@wiwi.uni-frankfurt.de

Michael Weiler
E-Mail: me@michaelweiler.de

Nora Jansen
E-Mail: jansen@wiwi.uni-frankfurt.de

Keywords Validation study · Contact diary · Digital footprint data · Personal network · Social capital · Convergent validity

1 Introduction

Nowadays, “social media networks” (hereafter “SMN”) (Kane et al. 2014, p. 275) such as Facebook (hereafter “FB”) have become an indispensable part of people’s daily routine (Kosinski et al. 2015). They help members maintain and strengthen existing relationships, establish new ones, or even find romantic partners, mitigating many spatial or temporal constraints (Ellison and Vitak 2015). In addition, noting declining participation rates in surveys and rising costs to administer them (Couper 2017), scholars often perceive SMNs as a panacea for social research, able to replace or complement survey-generated data (Schober et al. 2016; see also Golder and Macy 2014). SMN users produce vast amounts of relational digital footprint data (Golder and Macy 2014) (also often referred to as digital trace data), which broadly speaking pertains to “records of activity (trace data) undertaken through an online information system (thus, digital)” (Howison et al. 2011, p. 769). Generally speaking, such digital footprint data are readily accessible at marginal costs, continuously generated in a nonreactive way without researcher intervention, and less prone to error due to self-reported biases (Schober et al. 2016). Thus, scholars might ask questions that previously were difficult to answer and these new data opportunities accordingly are fostering unprecedented empirical research on social capital (e.g., Hinz et al. 2015). In general, the theoretical lens of social capital has received much attention from many scholars who have provided evidence that it is a suitable foundation for studying all types of phenomena that are relevant to disciplines such as information systems (hereafter “IS”) (e.g., von Stetten et al. 2012; Cummings and Dennis 2018; Weiler et al. 2022) or sociology (e.g., Norbutas and Corten 2018; Lórinicz et al. 2019). Moreover, it is also a relevant and established perspective for business-related research (e.g., Florin et al. 2003; Li et al. 2013; Lüdeke and Allinger 2017).

According to Nahapiet and Ghoshal (1998), it is possible to define social capital as “[...] the sum of the actual and potential resources embedded within, available through, and derived from the network of relationships possessed by an individual or social unit. Social capital thus comprises both the network and the assets that may be mobilized through that network” (p. 243). Apart from this definition, numerous other ones exist (see Adler and Kwon 2002 for an overview), which have their roots in all kinds of academic disciplines, including sociology (e.g., Bourdieu 1986) and economics (e.g., Loury 1992). Despite their different scientific origins, the definitions nonetheless all share the underlying tenet that the embeddedness in certain social structures (i.e., networks) provides access to many types of resources that facilitate an individual actor’s actions and his or her goal attainment (Coleman 1988; Nahapiet and Ghoshal 1998; Adler and Kwon 2002 also for a detailed review of the social capital concept). Regardless of being an established concept, which is leveraged across many disciplines, the measurement of social capital presents itself nonetheless as a major challenge to researchers (see “Theoretical Background” section for more details). However, the omnipresent digitalization of everyday life

and its associated perpetual generation of digital footprint data by individuals using the corresponding digital platforms might help circumvent the measurement issue. This is because this new promising data source might provide a unique opportunity to extract indicators that mirror those obtained from established traditional data collection tools. Yet, the usefulness of indicators extracted from digital footprint data for testing theories involving social capital depends largely on whether they are sufficiently informative in relation to their counterparts measured with traditional social science methods, such as surveys (Bisbee and Larson 2017). Thus, the establishment of convergent validity with extant valid indicators is crucial for digital footprint ones, as relying solely on their face validity may result in unreliable conclusions (Jungherr 2018). In addition, Appel et al. (2014) pointed to the need of conducting validation studies of social capital metrics, as they demonstrated in their own research that popular and easy to implement psychometric scales of the concept cannot be considered as valid metrics of structural social capital. In the present paper, we aim to broaden this knowledge. Specifically, in this study, we explicitly seek to assess the convergent validity of the social capital construct (i.e., its structural dimension in particular) by comparing digital footprint data with an established, self-reported, structural measure. We thereby depart from earlier attempts (e.g., Johnson et al. 2012; Von Der Heide et al. 2014) by taking a hitherto unique methodologically route to elicit self-reported egocentric network data. In detail, we use a contact diary approach that offers detailed, longitudinal data about actual and active personal networks (Fu 2007; Dávid et al. 2016; Yen et al. 2016). Although it is a burdensome and difficult approach to measure social networks, it is superior to others, such as name generators, single-item questions, and sensors, as it is almost predestined to concurrently capture ties of various strengths in all kinds of social contexts (Fu 2007; Dávid et al. 2016). Thus, contact diary data is the perfect “[...] yardstick to evaluate the validity of other techniques” (Fu 2007, p. 213). In particular, we asked our participants to record every social contact they have made during the course of one week. To the best of our knowledge, we are the first who use this sophisticated method in order to explicitly assess the convergent validity of metrics extracted from digital footprint data in measuring the structural dimension of social capital.

The rest of the article is organized as follows. The next section elaborates further on the concept of social capital, focusing on its multidimensionality and its measurement. The third section summarizes extant studies related to our research purpose. The fourth section elaborates on our conceptualization of social capital as it is used in the present study. The fifth section describes our data collection and research methodology, while the sixth section details our findings. The article ends with a discussion of our results and their implications for the research community.

2 Theoretical Background

2.1 Social Capital—A Multidimensional Concept

Social capital is not only a multilevel (Payne et al. 2011) but also a multidimensional concept, comprising structural, relational, and cognitive dimensions (Nahapiet and Ghoshal 1998; Tsai and Ghoshal 1998). To put this in perspective, the structural dimension of social capital relates to the specific structure of ties between individuals, or in other words, the presence or absence of relationships between them. The relational dimension takes into account the quality of a relationship, as it focuses on aspects that include tie strength, trust, and obligations, as well as expectations. Meanwhile, the cognitive dimension highlights the meaning of a shared language and understanding between actors as necessary components for successfully exchanging the desired tangible or intangible resources (Nahapiet and Ghoshal 1998; Bolino et al. 2002). In the present study, we focused decidedly on the structural dimension of social capital, which is central, as nearly every definition of the concept agrees with this dimension (Adler and Kwon 2002). This is the case because the other two dimensions rely on the structural dimension to flourish. For instance, only if an actively cultivated tie between two individuals is present (i.e., structural dimension) trust (i.e., relational dimension) can grow in their relationship (Tsai and Ghoshal 1998). Besides, we also examine the concept of social capital from an individual-level perspective. Thus, we consider it as a private good, implying that a personal investment into social relations leads to exclusive beneficial returns, but only to the individual investor (i.e., the one who actually makes the effort to cultivate the corresponding ties) (Bourdieu 1986; Lin 1999). Proponents of the collective-level perspective, however, emphasize the public good aspect of the concept of social capital, which mirrors the idea that everyone in a community will benefit from the created social capital, not just those taking part in its immediate development (Coleman 1988; see also Lin 1999).

2.2 Measuring Structural Social Capital

In general, measuring social capital remains a major challenge for researchers though (Engbers et al. 2017), because it depends on the research context as well as the available data. A prominent operationalization of the structural dimension of social capital uses the node degree (e.g., Hinz et al. 2015), which refers to the number of direct contacts—that is, the size of the egocentric network. Using this metric to measure social capital requires the researcher to act under the assumption that having more contacts increases the chance that some of those contacts will provide the focal actor with necessary resources (Borgatti et al. 1998).

In the context of real-world networks, name generators or contact diaries are usually used to identify the total number of an individual's contacts (Fu 2007). Despite extensive validation of these measures, they also are laborious to collect and still subject to some limitations (Marsden 2011), such as self-report and recalling issues (e.g., Brashears et al. 2016) or interviewer effects (e.g., Herz and Petermann 2017). Yet measuring the structural dimension of social capital has become easier,

due to the inherently relational nature and ready availability of digital footprint data (Golder and Macy 2014). For instance, Dissanayake et al. (2015) harvested digital footprint data from the crowdsourcing platform Kaggle to measure this dimension of social capital. In addition, SMNs offer promising digital representations of real-world networks, further strengthening the argument that these data can accurately measure social capital and test related theories. For example, people use FB mainly to connect with previously known contacts (e.g., Ellison et al. 2007), which could lead to significant overlap between the data found at SMN and the real-world network of individuals (e.g., Van Zalk et al. 2014). Dunbar and colleagues (Dunbar et al. 2015, Dunbar 2016) indicate that the layered structure and size of the egocentric networks people maintain in SMNs mirror those of their real-world counterparts (see also Arnaboldi et al. 2013). This evidence suggests that while SMNs provide new ways to communicate and maintain networks, people still interact and behave similarly online and in the real-world (Arnaboldi et al. 2013; Dunbar et al. 2015). However, although these studies provide valuable insights, they also include only one type of data, either digital footprint data (Dunbar et al. 2015) or self-reported survey data (Dunbar 2016), to arrive at their findings about the structural properties of the online world. Subsequently, they use these gathered insights and validate them against established findings from previous studies about the properties of the real-world and therefore, lack the possibility of explicitly addressing the question of convergent validity of self-reported data and data extracted from digital footprints for measuring social capital¹.

3 Related Research

Prior research that addresses whether digital footprint data complement or substitute for self-reported data (e.g., Eagle et al. 2009; Kosinski et al. 2013; Von Der Heide et al. 2014; Mastrandrea et al. 2015; Socievole et al. 2016; Jungherr et al. 2017)² suggests ambiguous indications about the potential value of these data, whereby the level of congruence varies with the nature of the sources and the research purpose. These studies often come from disciplines such as computer science (e.g., Arnaboldi et al. 2013), psychology (e.g., Kosinski et al. 2013), political science (e.g., Jungherr et al. 2017), or epidemiology (e.g., Mastrandrea et al. 2015); therefore, they pursue other epistemological insights. Hence, it is not surprising that they measure real-world networks in terms of spatial proximity, as recorded by wearable sensors and phones (e.g., Kibanov et al. 2015; Socievole et al. 2016) or the number of daily face-to-face encounters (e.g., Gaito et al. 2012). At best though, they can provide only implicit insights into the validity of digital footprint data for measuring real-world social capital. Those few studies that use sociometric surveys to measure structural real-world social capital (e.g., Johnson et al. 2012; Von Der Heide et al. 2014) are also limited in a reliable and generalizable assessment of the convergent validity due

¹ For a detailed review of further studies that examine the relationship of online and offline social capital we refer to the online appendix.

² In the online appendix, in Table A1, we further describe, summarize, and tabulate the reviewed studies.

to their very small sizes ($n < 50$) and/or their focus on certain study populations such as employees of a specific organization. Thus, the strength of our study compared to those prior ones also lies in our larger sample of individuals, who come from all walks of lives (i.e., students, retirees, people working in all kinds of professions and organizations) (see the “Data and Method” section for more details). As a result, our heterogeneous sample of recruited individuals enables us to surmount the mentioned generalizability issue. Only one larger-scaled survey experiment (Bisbee and Larson 2017) explicitly highlights that it is an expedient option to exploit digital footprint data (defined as e-mail usage and/or SMNs membership) to study real-world social capital. However, the authors tend to focus on the tie strength (i.e., the relational dimension of social capital) and do not consider the node degree metric, a common measure of the structural dimension of social capital. Nor do they use actual digital footprint data to arrive at their conclusions.

4 Conceptualization of Social Capital

Before further elaborating on the topic why the assessment of convergent validity is crucial in testing theories, we distinguish between the terms “social media network-social capital” (hereafter “SMN-SC”) and “overall-social capital” (hereafter “O-SC”), as used in this study. The latter refers to the node degree measured with a contact diary method. O-SC considers all self-reported active contacts, including those that occur by phone, e-mail, video chat, SMN, text message, letter, or face-to-face interactions. The former (i.e., SMN-SC) refers to the same metric but extracted from digital footprint data, such as FB friendship data, which are a byproduct of the ubiquitous IS usage (Golder and Macy 2014). The notion of O-SC might signify that it encompasses all dimensions of the concept of social capital. However, it is important to emphasize at this point once more that in the way we use the term, we refer only to its structural dimension (i.e., the presence of social ties and the consequent network size). As just elaborated, O-SC, therefore, in our sense, reflects that every communication channel and way of interacting counts toward grasping the node degree of an individual, not just the digitized friendship processes happening in the SMN of Facebook (i.e., SMN-SC).

An IS creates a repository of all accumulated digitalized social relationships within the unique platform’s boundaries, which are steadily added to the contact list beginning on the day of the user’s registration. Thus, the digitally extracted metric might include not only actively maintained strong(er) and weak(er) ties but also ties that at one point crept into the user’s SMN contact list and then may have remained in a dormant state since. Dormant ties are always lying in wait to potentially be reactivated and then in turn provide access to (the needed) resources (Levin et al. 2011)³.

We depict and summarize our ideas in our conceptual model (see Fig. 1). We consider SMN-SC to be a subclass of O-SC and refer to this integrated point of

³ More details on the characteristics of SMN in social capital accumulation and maintenance processes are given in the online appendix.

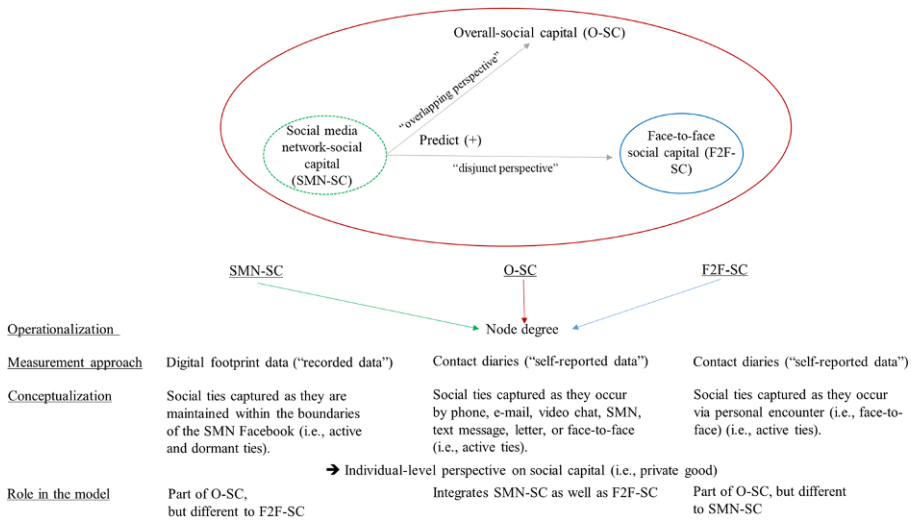


Fig. 1 Conceptualization of social capital in the present study

view as the “overlapping perspective”. Specifically, individuals’ use the available IS just as another convenient way to develop and groom their social relationships. For instance, two friends exchange several WhatsApp messages during the day, and then in the evening, they have an after-work beer together at their much-loved bar. At home, they then reflect on their meeting from earlier that day and tag each other on FB. These hypothetical interactions exemplify how individuals today use many different communication channels throughout the day, with each possessing their own unique characteristics, to stay in contact and to cultivate their relationships.

Figure 1 additionally shows that we can conceptualize the relationship between the digital and the real-world according to a more conservative point of view, which we refer to in this paper as the “disjunct perspective”. It suggests that both worlds are mutually exclusive empirical spheres that are also somehow simultaneously interrelated. The main argument behind this perspective is that “offline is the same as ‘real life’ and that online is ‘virtual’ and therefore less real” (Eklund 2015, p. 528). Thus, individuals maintain relationships in both spheres, but they tend to differ in terms, such as quality (e.g., Glüer and Lohaus 2016).

In addition, we can anticipate a positive link between O-SC and outcome measures, such as career success, as confirmed by various studies (e.g., Flap and Völker 2001). However, without further empirical knowledge regarding the convergent validity, we cannot conclude whether SMN-SC exerts the same effect on such outcomes. Two scenarios thus emerge with regard to using SMN-SC to proxy for O-SC. First, if SMN-SC is not similar to O-SC, indicators based on digital footprint data are unsuitable to test social capital theories, because we cannot anticipate the same effects on the outputs. Second, if we can infer O-SC from SMN-SC, we can reasonably assume that SMN-SC has the same or similar effects on outcomes, and therefore, it is appropriate to use SMN-SC as operationalization or proxy to test theories involving the social capital concept (Bisbee and Larson 2017).

5 Data and Method

Prior to the study period, January–May 2017, we developed two custom-built tools to collect data from two different sources: a website-based contact diary (or web diary, hereinafter) that was also optimized for the use on mobile devices and an application to harvest the non-public digital footprint data of the corresponding participants from FB's application programming interface (API).

5.1 Web-based Contact Diary Platform

We used validated questions from prior diary studies (e.g., Fu 2007; Dávid et al. 2016) to capture the wide-ranging content of the diaries. We instructed each participant (ego) to answer 21 questions about each newly recorded contact (alteri). The diary items encompassed (a) questions about the alteri's (sociodemographic) characteristics, (b) information about the specific contact situation like the location, and (c) questions about the participant's relationship to the alteri, such as emotional closeness (see also Dávid et al. 2016). Regardless of how often the participant had contact with a given network member within the period of investigation, they only needed to provide information about that member's characteristics or fixed features of their relationship, such as how much they like each other, once. For repeatedly encountered contacts, the participants only needed to answer the questions from block (b), resulting in three questions. To mitigate the workload, participants could also add contacts in advance into the database. For such not yet contacted individuals, but who the participants thought they will surely contact during the data collection phase, the participants only needed to answer questions from blocks (a) and (c).

As guidelines, we adapted and built on Fu (2007, p. 199) one-on-one contact definition: "Please record every *considerable* contact that you have made today *with all individuals you know by name, whereby it does not matter who initiated the contact*. These contacts can include all kinds of one-on-one contacts such as saying hello, chatting, talking, meeting, or sending or receiving a message, that occurred face-to-face, over the phone, on the Internet, or by other means of communication." To narrow the wide-ranging original definition, we added two restrictions. First, the participant must know the contacted individual by name (DiPrete et al. 2011), which excludes contacts with strangers but still includes relevant strong and weak ties. Second, the participant must regard the contact as considerable (see also Dávid et al. 2016), which should exclude contacts known inevitably by name but to whom the participant ascribes no lasting importance for his/her daily life, such as a conventional call with a customer service representative or a technician repairing the TV set. Encounters in the latter sense are somehow generic, as they are not linked to a specific individual but to the role this individual fulfills in the kind of transaction.

We instructed participants to record their contacts for one week, a typical period of investigation in diary studies (e.g., Dávid et al. 2016; see also Fu 2007). To capture the contacts accurately and mitigate recall issues, we encouraged participants to register each contact on the same day it occurred, though the web diary also offered an opportunity to enter contacts from the previous day, in case they forgot or had no time to do so on the day when the contact occurred. In addition, we asked them

to record the FB (nick)name of the contacted individual, which sometimes differs from their real name. With this step, we ensured that we could accurately link the recorded names in the contact diary with the corresponding names used on the IS⁴. Participants who had not recorded their contacts by 8:00pm received a reminder e-mail with an access link to the diary, to trigger contact recording for that day, which also helped reduce the non-completion rate.

After participants signed into the web diary with their personal FB account for the first time, we automatically assigned them a distinctive identifier, to ensure their anonymity but also support the ongoing combination of the data sets. During their first web diary login, an authentication screen popped up, asking for permission to collect their basic FB data such as name, birth year, gender, and friends list. Their initial consent was required to continue with the web diary; those who did not consent could not participate and were redirected back to the login page. Participants who granted us access to their basic FB data were led on to a one-time questionnaire that gathered sociodemographic and personal characteristics (see “Control variables and moderators (interaction terms)” subsection). After they answered these mandatory questions, they entered the actual contact diary page, where they could begin to record their contacts. During their next visit to the web diary, participants also had to give us explicit consent for an extended set of FB permissions, including access to their FB interaction data, such as posts, comments, messages, likes, and tags, as well as information about their preferences and activities (e.g., group memberships).

Connected to this, we emphasize that it is ethically and morally imperative to properly and consciously handle the highly sensitive data (i.e., personal (network) data) that the participants shared with us to protect their privacy. Thus, we implemented mechanisms to help ensure this crucial endeavor. First, we provided the participants with a detailed description of the study and its purpose to comprehensively inform them about the study and allow them to adequately assess whether they want to participate in the study or not. In addition, this evinced information led them to choose the kind of project in which they eventually participated. As mentioned, all the participants recruited voluntarily participated in our study and explicitly gave their consent and permission for the collection of their private digital FB footprint data as well as self-reported survey and contact diary data. In this regard, we emphasize that we only retrieved digital footprint data, which were expedient for our research purpose, thereby adhering to the principle of data minimalization. For instance, to protect the participants’ privacy, we only collected information about their number of FB interactions and did not retrieve or store any content related to these exchanges. While renouncing the storage of any material related to the content of the FB interactions, we also inherently respected the privacy of the participants’ contacts, as no immediate need arose to also obtain their consent to use such information. Second, naturally, given the highly sensitive nature of the collected information, we unambiguously assured our participants that the digital footprint data and self-reported survey and contact diary data they provided will only be used for our specific research purpose and will, under no account, be shared

⁴ We elaborate on this point in the online appendix.

with other parties. Informing the participants that their data will be kept confidential further promoted their trust in sharing their sensitive data with us⁵.

5.2 Sample

Given its global presence and immense number of active users from all walks of life, FB seems to be the natural choice of a SMN for our research purpose. To recruit the specific target population—people with a registered FB account—we took several steps. First, we applied a common approach in diary studies and recruited members of the researchers' social networks (Fu 2007). Second, we advertised the study during several lectures held at a major European university and approached pedestrians in public spaces. In total, we were able to recruit 181 participants through these efforts. However, not everybody of those eventually qualified for our sample: On the one hand, we had to exclude 26 (14.36%) participants, because they either only provided information to the questions asked in the one-time questionnaire and/or they only provided access to their digital footprints, but they did not start recording any contacts into the web-diary. On the other hand, we had to discard 24 (13.26%) further participants, as they did not meet the recommended inclusion criteria (see Ohly et al. 2010) to record their interactions for at least five days. Hence, 131 (72.38%) participants qualified for being part of our convenience sample. However, we had to remove three (2.29%) further participants, because they also did not comply with the given instructions, as they recorded only one contact over seven days. The final sample comprised information from 128 (97.71%) people, who received 20€ as compensation. We also included a self-monitoring feature that confirmed their continuous effort to record their contacts. With this self-designed feature, participants could view summarized statistics of their networks and observe their growth during the diary-keeping period, including the frequency of the used communication channels, (average) number of contacts/interactions, and name of the man and woman most often contacted in the previous seven days.

5.3 Measures of Social Media Network-social Capital and Overall-social Capital

In line with prior research (e.g., Hinz et al. 2015; Weiler et al. 2022), we used the node degree to operationalize the structural dimension of social capital. This metric, as a proxy for social capital, reflects the assumption that the number of direct contacts indicates the amount of accessible resources (Borgatti et al. 1998). While a bouquet of social capital measures exists, they differ in their potential to grasp the concept. Yet, according to Borgatti et al. (1998), the node degree is

⁵ In light of the European General Data Protection Regulation (GDPR), a new way of leveraging digital footprint data has emerged that inherently makes the consent of study participants a mandatory prerequisite for obtaining their data. Specifically, the GDPR equips SMN users with the agency to consult all their personal information that is stored by platform service providers, allowing platform members to request their data, usually with just a few clicks, and then to hand it over to scholars who can then utilize the data for research endeavors (Stier et al. 2020). For a more detailed discussion of this new data collection method see e.g., Boeschoten et al. (2020).

unambiguously associated with the concept and therefore emerges as an adequate and established measure for it: “The first set of measures, closest to the verbal description of social capital, consists of the standard ego-network measures that are well known in the network literature” (Borgatti et al. 1998, p. 30). Lőrincz and Németh (2022, p. 1124) reaffirm this conjecture while utilizing digital footprint data: “Several indicators were proposed to measure social capital using social network data, but there is a consensus over the simplest network indicator; the number of connections (degree).”

With the diary method, we gathered the dependent variable, O-SC, measured as all participants’ active, self-reported, direct contacts, regardless of how they happened. The digital footprint data provide the measure of the independent variable, SMN-SC, which pertains to the number of digitalized, direct FB friends (i.e., FB network size). Eventually, we log-transformed our independent and dependent social capital variable using the natural logarithm, in order to address the identified skewness and kurtosis concerns⁶. Examining the individual’s network size (i.e., the structural dimension of social capital), we inherently treated the ties that constitute the resulting social structure as equal, as we could not differentiate between their specific strength (i.e., strong or weak). However, it is important to remember that depending on the strength of the relationship (i.e., the relational dimension of social capital), different content and support acts are exchanged between the dyad members. For instance, strong ties are more likely to provide emotional support than weak ties (Krämer et al. 2021).

6 Results

6.1 Assessing Convergent Validity of Digital Footprint and Self-reported Measures

While the participants in our study had on average 297.15 FB friends (median=246.5; SD=196.08; min=9; max=1015), they cultivated an average of 24.38 (median=22; SD=12.14; min=6; max=66) contacts utilizing personal encounters and/or further (digitized) ways of interacting with each other (i.e., O-SC)⁷. Some studies reveal a positive correlation between digitalized and real-world social networks (e.g., Von Der Heide et al. 2014), but others do not replicate this association (e.g., Socievole et al. 2016). We confirm the latter studies, in that we find no significant Spearman’s rho correlation between the two structural social capital metrics ($r_s=0.0851$, n. s.) at first glance.

To gain confidence in our result, we ran several robustness checks. First, some studies suggest that active management of FB relationships is important to gain social capital benefits (e.g., Ellison et al. 2014). Therefore, we correlated O-SC with the

⁶ O-SC: skewness=1.296; kurtosis=4.754. SMN-SC: skewness=1.311; kurtosis=4.898. We also assessed the normality of the data using the Shapiro-Wilk test, which suggests that both variables are not normally distributed ($p<0.05$).

⁷ In the online appendix, we also provide some details on the overlap of O-SC and SMN-SC.

number of active FB communication partners. Similar to previous work (e.g., Dunbar et al. 2015), we define an active communication partner as someone who shares at least one registered digital FB footprint, such as one exchanged message, with the participant. Our study participants averaged 156.56 active FB communication partners ($SD=103.13$; median = 142.5; min = 2; max = 543), similar to the so-called Dunbar number, which indicates that humans can actively maintain 150 contacts concurrently (Dunbar 2018). The Spearman's rho correlation coefficient ($r_s=0.1218$, n. s.) reaffirms that we cannot proxy O-SC with SMN-SC straightforwardly.

Second, recognizing that some participants are members of other SMNs, we correlate O-SC with the number of contacts maintained on XING, a work-related network, which is considered to be the European version of LinkedIn. For 38 participants who were also members of the XING platform, we were able to retrieve their number of contacts manually by visiting their profile pages ($M=114.5$; $SD=145.07$; median = 63.5; min = 2; max = 626). Again, the insignificant moderate Spearman's rho correlation coefficient ($r_s=0.2234$, n. s.) suggests that a straightforward measure is not possible.

Finally, as previously outlined, we can view the relationship between the digital and the real-world according to two perspectives (see Fig. 1). Thus, in the following correlational analysis, we take a closer look at the perspective we posit to be the “dis-junct one”. As we highlighted in Fig. 1, F2F-SC, as we define it in our manuscript, refers to social ties that were constituted only via personal direct encounter (i.e., face-to-face) between the involved individuals in the corresponding contact. Thus, we utilize our detailed contact diary data to extract the personal network size of an individual consisting solely of her or his face-to-face met contacts, which reflect off-line social capital in its traditional sense. Specifically, to construct this measure, we had to exclude three participants from our analysis, as they reported that they had not contacted any of their friends and acquaintances personally (i.e., face-to-face) during the data collection phase. This data-cleaning step therefore reduced the number of valid cases to 125 for this specific analysis. Those participants in our study had on average 15.50 personal face-to-face encounters (median = 15; $SD=9.35$; min = 1; max = 54). We again performed Spearman's rho correlations to investigate the relationship between face-to-face social capital (hereafter “F2F-SC”) and the social capital metrics used above. O-SC correlated very strongly with the newly extracted variable F2F-SC ($r_s=0.8112$, $p<0.05$), indicating strong convergent validity. However, SMN-SC did not significantly correlate with F2F-SC ($r_s=0.0757$, n. s.). This insignificant correlation mirrors our previously established results. Likewise, we also identified an insignificant Spearman's rho correlation coefficient between F2F-SC and the number of contacts maintained on the work-related SMN XING ($r_s=0.0215$, n. s.). Once more, these analyses corroborate that a straightforward measure between metrics collected with self-reported data and extracted from digital footprints is not attainable.

6.2 Multivariate and Moderator Analysis

Differences exist in how individuals behave within the digital sphere, which may explain why we did not find a significant correlation coefficient via a simple bi-

variate analysis: Several studies have already been published on how SMN users' motives differ, resulting in varying usage patterns (e.g., Joinson 2008; Alhabash and Ma 2017). Such heterogeneous behaviors imply that different types of individual SMN usage also affect the accrual of digitalized social capital (e.g., Brandtzæg 2012). Often, these differences in behaviors manifest itself along sociodemographic variables, which is also the case in the digital sphere (e.g., Muscanell and Guadagno 2012). For instance, Krasnova et al. (2017) demonstrate that men and women have different intentions behind remaining a member of a SMN. Other studies point to variables, such as age (e.g., McAndrew and Jeong 2012) and personality (e.g., Liu and Campbell 2017), as being important factors affecting levels of SMN usage.

Hence, a rather shortsighted, straightforward examination might not be sufficient in identifying the true relationship between both spheres (i.e., digital and real-world), and a detailed investigation seems necessary to arrive at a more accurate and realistic depiction of social reality. In other words, we need to consider important control variables and moderators that reflect individual differences, while evaluating the convergent validity of indicators extracted from digital footprint data with an established, self-reported, structural measure, in order to arrive at a deeper understanding of the relationship between SMN-SC and O-SC. In the next subsection, we will elaborate on these moderators, as well as other important control variables, which we will account for in our ordinary least squares (OLS) regression analysis.

6.2.1 Control Variables and Moderators (Interaction Terms)

First, according to digital divide theory (Van Dijk 2013), people must acquire the technical skills to use FB; without the corresponding competencies, mobilizing social capital on such platforms is difficult. To measure these skills, we computed a composite measure of the two highest loading items from the social skills factor, identified in Van Deursen et al. (2016) Internet Skills framework ($M=4.656$, $SD=0.535$, Cronbach's $\alpha=0.7492$). These two items read as follows: "I know which information I should and should not share online" and "I know when I should and should not share information online".

Second, digital skills are only a prerequisite for participation; users also must use the systems in capital-enhancing ways (Van Dijk 2013), which should ultimately lead to the creation and maintenance of social capital. Prior research highlights that using FB for informational (Gil de Zúñiga et al. 2012) or communication (Junco 2013) purposes is positively associated with social capital benefits. We thus control for *different types of Facebook use* by adapting validated items from prior studies, such as Correa's (2016) distinction of social ("contact friends and acquaintances"; "chat with other users"; "read comments"), informational ("give opinions about politics and public affairs"; "put links to articles"), and mobilizing ("create FB pages"; "create or summon events") FB uses. We also extend social usage by including additional items, similar to those identified by Junco (2013): "posting/tagging photos" and "commenting on content created by FB friends." In addition, we extend informational usage with three items ("stay informed about current events and public affairs," "get news about current events from mainstream news media," and "get news about current events through friends"), as suggested by Gil de Zúñiga

et al. (2012). We include items that refer to entertainment usage too (“playing online games,” “get information about your hobby,” and “viewing videos/photos”). To measure all these items, we asked respondents, “How often do you use FB for [type of usage]” (five-point Likert scale, ranging from “Never” to “Daily”). To construct the formative composite measure, we performed an exploratory factor analysis with all 16 items to identify the underlying variable structure and reduce data complexity. We stepwise-omitted items with factor loadings below the threshold of 0.45 (see Tabachnick and Fidell 2007): “create FB pages,” “playing online games,” “tagging photos,” and “create or summon events.” The resulting factor solution comprises 12 items related to the social, informational, and entertainment types of FB usage. The remaining items all had loadings greater than 0.45 on the first factor and explained a substantial amount of variance (72%). Based on this factor analysis, we constructed a composite measure of types of FB usage ($M=2.589$, $SD=0.755$, Cronbach’s $\alpha=0.8713$).

Third, we control for extraversion, because existing research shows that this personality trait strongly influences real-world network size (see Selden and Goodie 2018 for a review) and SMN size (e.g., Chen 2014). We use a validated, abridged version of the Big Five Inventory, as adopted by the German Socio-Economic Panel Study (Hahn et al. 2012), to measure the mentioned personality trait ($M=5.033$, $SD=1.139$, Cronbach’s $\alpha=0.7654$). Participants rated their agreement with items on a seven-point Likert scale (ranging from 1 = “Does not apply at all” to 7 = “Totally applies”). An example item reads as follows: “I see myself as someone who is communicative, talkative”.

Fourth, previous studies highlight that standard sociodemographic variables like gender, age, and education affect social network sizes too (e.g., Kalmijn 2012; Brashears et al. 2016). Participants provided information about their sociodemographic (age and gender) and socioeconomic (secondary school graduation and vocational training) backgrounds. We measure these variables according to standards suggested by the German Federal Statistical Office (Statistisches Bundesamt 2016). In our study, 56.25% of the 128 participants were women (43.75% men). The average age of our participants was 32.5 years (median = 28; $SD=11.77$; min = 16; max = 68). We use the two variables, “secondary school graduation” and “vocational training,” to classify the participants’ level of education, in line with the International Standard Classification of Education (OECD 2015). Then we aggregate this information to three levels of educational attainment: low (0.78%), middle (43.75%), and high (55.47%).⁸ In the convenience sample, only one participant indicated a low level of education, so we combined the low and middle levels for further analysis.

In addition, noting research that shows that gender regularly interacts with other variables (e.g., Skiera et al. 2015), producing specific results for men and women, we include three two-way interaction terms; a pooled estimation might obfuscate such relationships (Skiera et al. 2015). The lower-order terms combine gender and level of education, gender and age, and gender and SMN-SC. Prior studies also identify age as a strong predictor of digitalized (e.g., McAndrew and Jeong 2012)

⁸ See [http://ec.europa.eu/eurostat/statistics-explained/index.php/International_Standard_Classification_of_Education_\(ISCED\)](http://ec.europa.eu/eurostat/statistics-explained/index.php/International_Standard_Classification_of_Education_(ISCED)) (accessed: 25.05.2017).

and real-world (e.g., Kalmijn 2012) network size. Thus, we also include a two-way interaction of age and SMN-SC in our model. Gender and age could both simultaneously moderate the effect of SMN-SC on O-SC (see also McDonald and Mair 2010), so we add a higher-order, three-way interaction term (gender×age×SMN-SC) to check systematically for its potential effects on our dependent variable.

6.3 Overall-social Capital Prediction

Table 1 contains the OLS regression estimates, which we computed with robust standard errors. Model 1 includes only the independent variable, SMN-SC. Model 2 adds the control variables. Their overall model fits are insignificant, so we do not detail these models further. These results corroborate our correlation analysis; we cannot directly measure O-SC with SMN-SC. In contrast, the more complex Model 3, which includes the control variables plus interaction terms, is highly significant ($F_{(12, 115)} = 4.57$; $p < 0.001$) and has a moderate variance explanatory value of 15.3%. The overall variance explained by the model increases by 7.66 percentage points when we include the interaction terms ($\Delta F_{(5, 115)} = 3.11$, $p < 0.05$). Taking both the control variables and the interaction terms into account reveals a relatively strong, significant correlation between SMN-SC and O-SC ($B = 0.446$; $p < 0.05$). This result signals convergent validity and implies that SMN-SC is positively linked to O-SC, so we can use both measures as direct substitutes, as long as we control for other factors and the interactions between them. For example, a 15% increase in SMN-SC is accompanied by a 6.4% increase in O-SC ($1.15^{0.446}$)⁹. To put our findings into perspective, we also estimated both a control-only model (Model 2a) and Model 3 while eliding the three-way interaction term (Model 3a). A brief comparison reveals that Model 3 has not only a much better model fit than Model 3a but once more also buttresses the finding that only after the inclusion of the three-way interaction term does the strong and significant correlation between SMN-SC and O-SC actually emerge.

Age relates to O-SC at a 10% significance level ($B = 0.066$; $p < 0.1$); older participants exhibit more O-SC (see also Hill and Dunbar 2003). We explain this finding by noting that most people naturally accumulate social capital during the course of their lives, due to increased embeddedness in work-related networks and voluntary organizations (McDonald and Mair 2010), which provide valuable opportunities to cultivate ties (Mollenhorst et al. 2014).

Because all our tested interaction effects are at least marginally significant, the effect of SMN-SC on O-SC varies along the different states of the moderators, which implies no clear connection of the social capital indicators. That is, the identified convergent validity strengthens or weakens depending on individual sociodemographic factors. The interaction term between gender and SMN-SC is significant

⁹ This parameter translated into words means the following: For example, if we want to see what difference a 15% increase in our independent variable (i.e., SMC-SC) will bring about in our dependent variable (i.e., O-SC), we need to compute 1.15 to the power of the coefficient of the conditional main effect (0.446). Subsequently, to transform the computed value into a percentage value, we subtract 1 from the computed value and multiply it by 100. Based on the just outlined, the general formula for obtaining the parameter reads as follows: $(1. \times ^{\text{regression coefficient}} - 1) * 100$.

Table 1 OLS regression models predicting overall-social capital (log)

	Model 1	Model 2: Controls	Model 2a: Without SMN-SC (controls-only)	Model 3: Controls & Interactions	Model 3a: Model 3 minus three-way interaction term
Social Media Network-Social Capital (SMN-SC) (log)	0.065 (0.057)	0.047 (0.069)	–	0.446* (0.219)	0.112 (0.161)
Male (Ref: Female)	–	–0.147 ^a (0.084)	–0.148 ^a (0.085)	2.411 (1.685)	–0.882 (0.766)
High Level of Education (Ref.: Low/Middle)	–	0.154 ^a (0.090)	0.146 ^a (0.087)	0.018 (0.124)	–0.003 (0.123)
Age	–	0.001 (0.004)	0.000 (0.004)	0.066 ^a (0.034)	0.014 (0.025)
Extraversion	–	0.051 (0.041)	0.058 (0.038)	0.064 (0.043)	0.059 (0.043)
Internet Skills (Social Skills)	–	0.009 (0.067)	0.008 (0.065)	0.006 (0.062)	–0.008 (0.064)
Types of Facebook Usage	–	0.042 (0.063)	0.057 (0.055)	0.032 (0.058)	0.048 (0.060)
Male (Ref.: Female) × Age	–	–	–	–0.085* (0.038)	0.003 (0.007)
Age × SMN-SC (log)	–	–	–	–0.013* (0.006)	–0.003 (0.004)
Male (Ref.: Female) × High Level of Education (Ref.: Low/Middle)	–	–	–	0.348* (0.179)	0.390* (0.176)
Male (Ref.: Female) × SMN-SC (log)	–	–	–	–0.532 ^a (0.292)	0.077 (0.123)
Male (Ref.: Female) × Age × SMN-SC (log)	–	–	–	0.017* (0.007)	–
Constant	2.726** (0.320)	2.354** (0.468)	2.574*** (0.324)	0.351 (1.279)	2.205* (0.963)
R ²	0.011	0.076	0.072	0.153	0.119
Δ Change in R ²	–	0.0657	–	0.0766*	–
F	1.296	1.541	1.735	4.573	2.104
N	128	128	128	128	128

Note: Unstandardized regression estimates; robust standard errors are in parentheses

^a $p < 0.1$

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$

Table 2 Typology of the relationship between social media network-social capital and overall-social capital

		Degree of overall-social capital	
		Smaller	Larger
Degree of social media network-social capital	Smaller	Lone Wolves	Traditionalists
	Larger	Onliners	Jack of all trades

at a 10% level ($B = -0.532$; $p < 0.1$), so the impact of SMN-SC on O-SC appears moderated by gender. As the node degree on FB increases, the gap between men and women in their predicted degree of O-SC also increases, and men who cultivate more SMN-SC indicate a smaller amount of O-SC than women do.

The interaction between gender and education level also is marginally significant ($B = 0.348$; $p < 0.1$). More educated men have greater degrees of O-SC than their less educated counterparts or women, regardless of their education level. A possible explanation for this finding is available in the substantial scholarship that examines how meeting opportunities affect the emergence of social relations (e.g., Mollenhorst et al. 2014). That is, more highly educated people have larger networks, because they participate in a wider variety of activities like cultural events (e.g., Lizardo and Skiles 2012). However, why does O-SC increase for highly educated men, but not for their female counterparts? Social role theory (Eagly et al. 2000) provides a possible explanation: Men and women become socialized into different roles while growing up, which in turn are linked to specific expectations, acquisitions of certain skill sets, and specific behaviors. Whereas the male role conventionally is less informed by emotions and more focused on ambitiousness, instrumentality, or goal orientation, the female role encompasses behaviors informed by helping others, sensitivity, relationality, and emotionality (Eagly et al. 2000). Thus, women tend to invest more time in existing relationships, which limits their time available to establish new relationships (Skiera et al. 2015), whereas men focus on the creation of new ones (Muscanell and Guadagno 2012) and nurture existing ties less intensively (see Dunbar 2018 for a review). Ties groomed by women are more intimate (Umberson et al. 1996), even in the work context (McDonald and Mair 2010), and mostly involve other women (Roberts et al. 2008), such that they are more emotionally intense than mixed-gender ties or man-man contacts (Roberts et al. 2009). However, limited time also constrains the active maintenance of a large number of emotional close ties (Roberts et al. 2009; see Dunbar 2018 for a review), as the maintenance becomes costlier as the ties grow more intimate.

We find a significant interaction effect between gender and age too ($B = -0.085$; $p < 0.05$), which corroborates previous empirical evidence that aging men confront decreasing network sizes (Kalmijn 2012). The significant interaction term indicates that for the male participants in our sample their predicted degree of O-SC decreases the older they are, whereas for their female counterparts it increases the older they are. In addition, we find a significant interaction effect between age and SMN-SC ($B = -0.013$). The negative two-way interaction term is significant at the 5% level, such that as SMN-SC increases, the gap between younger and older people's degree

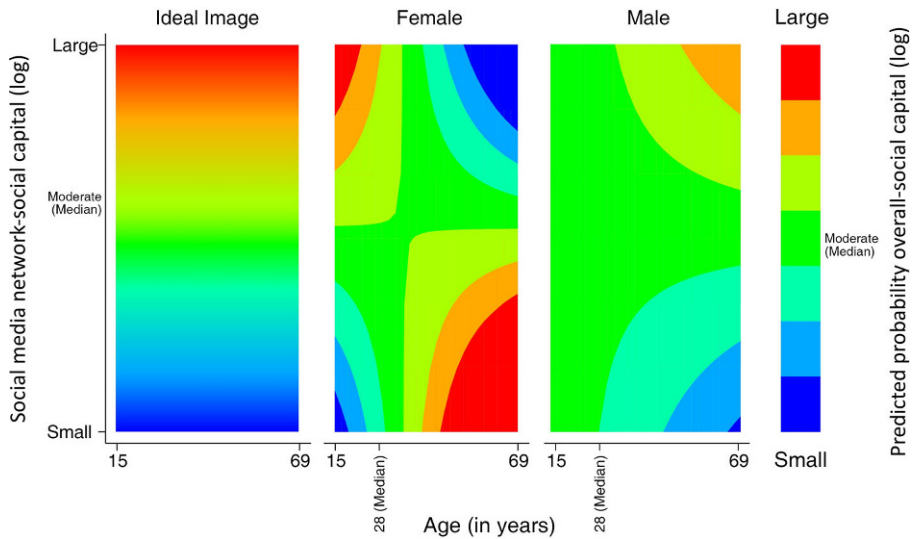


Fig. 2 Predicted probability of overall-social capital (log) by age and social media network-social capital (log)

of O-SC increases. Older people who cultivate more SMN-SC exhibit less O-SC than younger ones do.

Finally, the estimated model indicates that the three-way-interaction of age, gender, and SMN-SC is significant ($B=0.017$; $p<0.05$). Because interpreting higher-order interactions with continuous variables is challenging, we take several steps to facilitate the interpretation. First, we define a typology of different nonexclusive combinations of how SMN-SC and O-SC relate. The resulting two-by-two matrix with four segments is in Table 2. We classify the types by the median value of SMN-SC (median: 5.507) and O-SC (median: 3.091), such that a smaller degree of social capital implies a logarithmic value below the median, and a larger degree means the logarithmic value is equal to or greater than the median.

Second, we use Stata's margins command to understand and graphically represent the three-way interaction. This command calculates predictions of the dependent variable, for which some or all independent variables are fixed at certain values (Williams 2012). In our case, by computing the margins we could determine how the predicted probability of having a certain degree of O-SC varies with age, level of SMN-SC, and gender (see Fig. 2). Again, we use a median split to categorize the participants as "young" (<28 years) or "old" (≥ 28 years) FB users, whereby we place those participants with exactly the median value of 28 years into the "older category". The y-axis depicts the logarithmized degree of SMN-SC, and the x-axis indicates respondent age. For the three-way interaction, we present two identical graphs, one for each gender. The logarithmized degree of O-SC is indicated by a color spectrum; generally, the warmer the color, the larger the amount of the predicted O-SC. As Fig. 2 shows, a woman younger than 28 years who has a smaller degree of SMN-SC also can be predicted to exhibit a smaller degree of O-SC (as

Table 3 Convergent validity between social media network-social capital and predicted amount of over-all-social capital

		Females	Males
Age (in years)	Degree of social media network-social capital		
<28	Smaller (less than 246 Facebook contacts)	✓ (0.80)	(✓) (0.07)
	Larger (equal and more than 246 Facebook contacts)	✓ (0.77)	✓ (0.37)
≥ 28	Smaller	– (–0.05)	✓ (0.38)
	Larger	– (–0.37)	✓ (0.31)

Note: Spearman's correlation coefficients in parentheses

indicated by cold, bluish colors), in support of convergent validity. The first graph in Fig. 2 depicts how the gradient would look with perfect convergent validity.

Third, to assess validity statistically, we use the predicted values of O-SC from our fitted regression model and calculate Spearman's rho correlation coefficients with the SMN-SC values. For descriptive purposes and to support the practical application of our results, we use an exponential function to transform the logarithmic median value of SMN-SC back to its original value ($\exp(5.50736)$). Thus, a smaller degree of SMN-SC for participants implies a node degree less than 246, and a larger degree refers to participants with a node degree equal to or greater than 246.

Before we address ourselves by discussing the findings, we would like to put our upcoming interpretation of them in some context. In general, some ambiguity exists regarding when to deem a correlation between two measures as sufficiently strong enough to establish convergent validity. Often, scholars consider a moderate correlation coefficient as appropriate (e.g., Appel et al. (2014) utilize a threshold of at least 0.31 while looking at the convergent validity of social capital measures). Against that backdrop, and given that the concept of social capital is inherently difficult to measure, we share such a perspective in the present manuscript and also consider a more lenient approach with the aforementioned threshold as reasonable to define a valuable proxy measure. We believe this should also help propel new insights into the concept of social capital. However, we acknowledge that researchers pursue different objectives with their research and normally have unique data sets at hand. Thus, some might be more comfortable with applying a more conservative cut-off value, only regarding larger positive correlation coefficients as suitable indicators for the identification of proxy measures (for what this perspective would imply for the interpretation of our study's findings, please see our fifth limitation in the "Limitations" section).

As Table 3 shows, it is possible to measure O-SC using digital footprint data from men with high SMN-SC (Jack of All Trades), regardless of age ($r_s > 0.30$). We also have confidence that it is valid to harvest digital footprints from younger women (<28 years) with smaller (Lone Wolves) and larger (Jack of All Trades) amounts of SMN-SC, in order to measure O-SC, because both metrics overlap to a substantial degree too ($r_s > 0.75$). The relationship also works for older men (≥ 28 years) with lower SMN-SC; we affirm a moderate significant correlation between the two measures ($r_s = 0.38$), which suggests their interchangeability. However, this potential is not universally applicable. Scholars should not treat harvested digital footprints

from young men with fewer than 246 FB contacts as valid representations of their O-SC, because the determined correlation coefficient is simply too weak to arrive at plausible conclusions ($r_s = 0.07$).

Moreover, we can classify women who are 28 years of age or older, depending on their degree of SMN-SC, as either “Traditionalists” or “Onliners,” which implies that it is not valid to use SMN-SC as a proxy measure for their O-SC counterpart. Again, social role theory provides a possible explanation for why the indicator validation does not work for older women categorized as Traditionalists. As noted, due to societal role expectations, women invest more time in cultivating interpersonal relationships than men do (Muscanell and Guadagno 2012; Skiera et al. 2015). Because men according to their social role invest less time in nurturing relationships than females do, they lose more ties as they get older, which leads to smaller real-world networks (Kalmijn 2012). Our data provide empirical evidence that women behave according to social role expectations and invest more time in maintaining social ties. In a t-test, using an “activity” variable obtained from FB interaction data, we test this claim. The variable measures how much time has passed between two FB timeline posts ($M = 230.258$; $SD = 831.98$), such that smaller values represent more activity on the FB timeline. The rationale behind this measure is that we can regard such publicly visible signals on timelines, such as comments, as investments that support the maintenance of relationships (Ellison et al. 2014). In turn, we find that women 28 years and older with smaller degrees of SMN-SC ($M = 93.797$; $SD = 112.821$) exhibit significantly more activity on their timeline than their male counterparts ($M = 570.207$, $SD = 1008.055$, $t(40) = -2.204$, $p < 0.05$).

FB users have various motives to expand their digitized networks. Some use the platform for socializing to increase their social capital, others, especially individuals who have an introverted personality, have low levels of self-esteem, or experience feelings of loneliness, balance these issues out by utilizing the anonymity and possibilities of SMNs to interact with other individuals. This latter behavior is known in the literature as the social compensation hypothesis (“poor-gets-richer”) (e.g., Zywicka and Danowski 2008). In accordance with this hypothesis, we find in a t-test that, compared with men ($M = 0.544$; $SD = 0.171$), women older than 27 years who exhibit a larger degree of SMN-SC ($M = 0.442$; $SD = 0.122$) have less overlap between their SMN-SC and O-SC ($t(28) = -1.831$, $p < 0.0778$). This indicates that these female “Onliners” possess weaker ties in their FB networks (Lankton et al. 2017). Thus, some women in our sample appear to turn to FB to mobilize support or feel socially connected, because they have difficulty cultivating real-world contacts, perhaps due to their health limitations (e.g., Oh et al. 2013), lack of mobility, geographically scattered contacts (e.g., Quan-Haase et al. 2017), low self-esteem/introversion (e.g., Zywicka and Danowski 2008), or loneliness (e.g., Böger et al. 2017). They also might befriend more others on FB because they miss potential opportunity structures to connect in the real-world, due to life-course transitions like child-rearing responsibilities (e.g., Munch et al. 1997), divorce (e.g., Kalmijn 2012),

Table 4 OLS regression predicting face-to-face social capital (log)

	Model 1	Model 2: Controls	Model 3: Controls & Interactions
Social Media Network-Social Capital (SMN-SC) (log)	0.036 (0.080)	0.096 (0.095)	0.996** (0.360)
Male (Ref.: Female)	–	–0.188 (0.140)	4.427 ^a (2.643)
High Level of Education (Ref.: Low/Middle)	–	–0.015 (0.150)	–0.212 (0.200)
Age	–	0.001 (0.008)	0.135* (0.052)
Extraversion	–	–0.034 (0.071)	–0.025 (0.073)
Internet Skills (Social Media)	–	0.248 (0.196)	0.236 (0.193)
Types of Facebook Usage	–	–0.190 (0.125)	–0.200 ^a (0.121)
Male (Ref.:Female)× Age	–	–	–0.137* (0.058)
Age× SMN-SC (log)	–	–	–0.026** (0.010)
Male (Ref.:Female)× High Level of Education (Ref.: Low/Middle)	–	–	0.511 (0.314)
Male (Ref.: Female)× SMN-SC (log)	–	–	–0.966* (0.465)
Male (Ref.: Female)× Age× SMN-SC (log)	–	–	0.027* (0.011)
Constant	2.309** (0.436)	1.544 (1.069)	–3.004 (2.212)
R ²	0.001	0.061	0.153
Δ Change in R ²	–	0.06	0.092**
F	0.210	0.922	2.149
N	125	125	125

Note: Unstandardized regression estimates; robust standard errors in parentheses

^a $p < 0.1$

* $p < 0.05$

** $p < 0.01$

or widowhood (e.g., Cornwell et al. 2008). Further research is needed to confirm this explanation¹⁰.

6.4 Robustness Check: Alternative Operationalization of Self-reported Social Capital

We also estimated a series of OLS regressions using our F2F-SC measure as our dependent variable to establish the multivariate relationship between our social capital metrics representing the “disjunct perspective” (see Fig. 1). Replicating our previous results, Table 4 highlights that SMN-SC is also significantly related to F2F-SC, but only after considering the moderating effects. The conditional main effect only turns significant if we control for certain factors and the interactions between them. Thus, the relationship between SMN-SC and F2F-SC is likewise strengthened or weakened depending on the pivotal, previously identified three-way interaction (gender× age× SMN-SC).

¹⁰ To gain confidence in our findings, we checked them for robustness by applying alternative regression model specifications. The detailed results of the robustness check are available in the online appendix.

Broadly speaking, our finding implies that these two perspectives (overlapping and disjunct) do not substantially differ from each other, despite their different conceptions and varying operationalizations of real-world social capital. In other words, the conceptual differences between both spheres, as delineated in Fig. 1, no longer seem to hold empirically.

7 Discussion

Typically, scholars apply network generators to capture the self-reported amount of structural social capital. Although they are established structural measures, the elicited networks are limited in size and error prone due to issues like recalling. Because the omnipresent SMNs, such as FB, store unprecedented amounts of digitalized social behavior, relational data extracted from these platforms might offer great potential for facilitating measures of the inherently difficult-to-grasp concept of social capital. Against that backdrop, in the present study, we aim to make a meaningful assertion about the potential of digital footprint data in facilitating the measurement of the structural dimension of social capital. The promise of digital footprint data has already been fathomed in other domains, for instance in a business-related context in terms of identifying opinion leaders (e.g., Jansen and Hinz 2022). We take a unique approach to examine the convergent validity of a structural social capital metric, node degree, measured with self-reported contact diary and digital footprint data. Without accounting for potential moderators, the bivariate correlation suggests that the metrics are not direct substitutes (see also Socievole et al. 2016). However, if we take a more fine-grained look at the data and estimate models that control for several potential moderating factors and thus include interaction terms, the insignificant relationship becomes statistically significant. Consistent with prior studies, we corroborate some conjectures of social capital theory, which suggests that digitalized and real-world social capital are related (e.g., Dunbar et al. 2015; Bisbee and Larson 2017).

Our results indicate a significant, conditional main effect in the presence of significant moderating interactions, including a three-way interaction term. This interaction term strengthens or weakens the convergent validity between SMN-SC and O-SC. SMN-SC collected from younger women and men (<28 years) seems particularly valid as a proxy for O-SC—in other words, people with a high FB adoption rate (see Blank and Lutz 2017; Greenwood et al. 2016) who became socialized without experiencing any divide between digital and real-world spheres. The relationship between SMN-SC and O-SC is stronger for younger women and men, but only if the latter exhibit more SMN-SC (i.e., at least 246 FB friends). For older people (≥ 28 years), we find a more nuanced picture. They began using SMNs when FB was just emerging as a serious competitor in the European market; other platforms like StudiVZ (Germany) or Hyves (Netherlands) were still very popular among the back then younger users. The association between SMN-SC and O-SC gets strengthened for men older than 27 years but weakened for women over this age. In the latter case, the amount of SMN-SC cannot be used as proxy for O-SC. Therefore, we do not advise testing theories involving the structural dimension of social capital

with the easily collectable digital structural social capital indicators harvested from women over the age of 27 years.

Researchers should not ignore SMN-SC as a practical alternative measure of O-SC per se; it provides a sound approximation when controlling for certain influential factors and the interactions between them. Nonetheless, the relationship between SMN-SC and O-SC is conditional on both age and gender, so it cannot be a homogeneous proxy for everyone. In summary, the convergent validity of both structural social capital measures seems particularly promising for younger women (<28 years) and older men (≥ 28 years), regardless of their amount of SMN-SC, as well as for younger men (<28 years) with more SMN-SC (i.e., at least 246 FB friends).

7.1 Implications

We contribute to the discussion about measuring the structural dimension of social capital by showing that the association between two metrics of the concept is ambiguous and more complex than anticipated. Specifically, we observe a bivariate relationship between SMN-SC and O-SC after controlling for several potential influential factors. The convergent validity between SMN-SC and O-SC depends, however, strongly on individual sociodemographic characteristics. In this sense, both scenarios we noted in the introduction regarding the effect of SMN-SC on outcome variables can occur simultaneously, and thinking solely in absolute terms about the use of SMN-SC to measure the concept is insufficient. Researchers need a refined understanding of the extent to which SMN-SC is informative of O-SC. Thus, our study adds new insights about the gendered nature of social capital and how it differs with age (e.g., McDonald and Mair 2010; Kalmijn 2012). For example, age appears linked to an increasing amount of O-SC.

We arrived at these findings because we recruited a heterogeneous sample of participants from all walks of life and therefore have obtained the data on their social networks and their sociodemographics. We thereby extended previous studies that recruited only certain populations such as students (see Table A1 in the online appendix for an overview). For instance, Stopczynski et al. (2014) gauged the interaction patterns of students across multiple communication channels and vetted how the resulting networks relate to each other. They strikingly demonstrate that the extracted networks from the different channels usually overlap to some degree, but that they are not entirely congruent. Each network provides supplementary information to the overall picture. However, given their focus on students, who are usually quite homogenous regarding their ages, the authors cannot leverage this essential variable to identify potential age differences. Like the present research, the study by Stopczynski et al. (2014) provided evidence that reminds us of yet another fact. Different methods can be used to collect the node degree (i.e., the structural dimension of social capital) and each method has its own inherent strengths and weaknesses that affect the resulting measures. Therefore, it is imperative to understand them and gauge how they relate to each other to install them with confidence as proxy measures.

In terms of practicality, our results reveal that it also is not sufficient to harvest just the degree of SMN-SC and start testing hypotheses—however tempting it might be. Instead, researchers must collect and consider additional factors before they can use SMN-SC as a proxy for O-SC. Using nonconverging proxies will lead to inconclusive, ambiguous empirical findings and ultimately result in fallacious theoretical implications (Carlson and Herdman 2012). With that insight, this study also extends knowledge of the relationship between digital and real-world social capital (e.g., Von Der Heide et al. 2014) by identifying the joint moderating roles of age and gender. As far as we know, no prior study has specified that the relationship between these measures is conditional on the interaction of individuals' age and gender. However, our findings indicate that this three-way interaction is critical to accurately gauge social reality. Yet, the effect of sociodemographic variables like gender and age on social network size also is already well established (e.g., Kalmijn 2012).

Despite our promising findings, researchers who prefer traditional measures of structural social capital might be reluctant to embrace the potential of digital footprint data if they do not agree that every tie maintained in the real-world would have the same counterpart in the digital sphere¹¹. Yet, these researchers have to keep in mind that what matters more for testing theories involving the concept of social capital than the degree of overlap is that both network structures mirror each other in terms of their size (see also Bisbee and Larson 2017). As mentioned previously, the node degree, as a structural measure of social capital, which we focus on solely in the present manuscript, captures only the potential access to resources, which are, however, not further specifically broken down to an explicit tie. Thus, if the structural metric of social capital (i.e., node degree) correlates, we can assume that individuals will have access to the important resources some way or another, even if the specific ties do not match one-to-one. In other words, “[...] many alters will give access to the same resources, and although similar resources available from several alters could be seen as a form of help ‘insurance’, usually one alter suffices to solve a certain problem” (Van Der Gaag and Snijders 2005, p. 3).

When the relationship between O-SC and SMN-SC is weakened and neither measure provides a direct substitute (e.g., for women 28 years of age and older), the combined use of self-reported and digital footprint data (hybrid design) seems suitable. Such an approach would help overcome biases associated with one sort of data (Jungherr 2018) but also result in more reliable measures of the overall degree and quality of an individuals' social capital. In their recent paper, Stier et al. (2020) address some pitfalls that emerge when combining both sources of data.

Digital footprint data are not a panacea or answer to all possible questions about social capital, which “still requires careful thought, rich substantive knowledge, and a relevant outcome measure” (Bisbee and Larson 2017, p. 520). Accordingly, the availability of vast digital footprint data should not be taken to mean that it would necessarily be appropriate to link digitized social capital indicators to any arbitrary outcome measure. Instead, the relationship between variables must be grounded in a sound theoretical foundation, not “ex-post rationalization” (Jungherr et al. 2017,

¹¹ This is also visualized by the partial overlap in Figure A1, see in the online appendix.

p. 352); random, highly significant relationships between variables in large data sets are not difficult to find (Jungherr 2018).

7.2 Business Implications

We can also distill some actionable insights for businesses from our findings. Specifically, previous research has shown that individuals with larger social networks are more influential and therefore more suitable for diffusing information in general (e.g., Katz and Lazarsfeld 1955) and for disseminating marketing campaigns in particular (Hinz et al. 2011). By using traditional methods of social network research, obtaining relevant information to identify who occupies such a strategically important position is usually ornate and only limited to a small sample of participants. Thus, marketers might leverage the node degree extracted from digital footprint data as valuable information to identify individuals who maintain a large social network and therefore might be especially suitable to start touting for a company's product. While our findings lend empirical support to this conjecture, our study also helps to better understand that the relationship between digitized and real-world social capital depends strongly on the sociodemographic characteristics of individuals. In order to reach individuals either obtaining large real-world or digitized networks, marketers can use our results to manage their marketing campaigns depending on whether they want the information about their products and services rather to be disseminated online or offline. Thereby, our findings are suitable for a microtargeting approach. For example, the age of male and female individuals indicates whether they possess a large degree of O-SC or not. For males who maintain a large digitized social network this is even independent of their age. This is also in line with previous literature, for example in terms of revealing opinion leaders online: Jansen and Hinz (2022) highlighted that a large number of online social contacts is a good indicator to identify male opinion leaders. In terms of female users who are older than 27 years with smaller degrees of SMN-SC, the study indicates that they tend to be more active on FB than their male counterparts. Thus, company marketers should keep this in mind and should fathom the marketing strategies of their various products and services according to the specific target audiences. Moreover, marketers should not expect that it is warrant that only while individuals maintain a large digital network that they can mobilize their corresponding real-world network in the same way to spread a positive message about the company and its products, as the latter mentioned social structure might lack the size of the former (and vice versa) (recall the typology of the relationship between SMN-SC and O-SC as depicted in Table 2).

Of course, not only the business subfield of marketing but also the domain of business research in general can be stimulated by leveraging social network-related measures extracted from digital footprint data. Specifically, we have to keep in mind that these observable data, as recorded on SMNs, not only originate from individual persons but also from business entities. This opens up unprecedented new avenues for future research. For instance, does the SMN-SC of a company's founder or CEO correlate with the social capital of the business, which this entity has on an IS? Put differently, do businesses whose founders maintain large numbers

of social ties also have a large following of individuals (potentially called business social media network-social capital), and does this for instance translate into the attraction of better-skilled employees (Saxton and Guo 2020)? Moreover, taking such a holistic perspective (i.e., combining the individual and organizational level) might also promise the development of new theories and the redefinition of existing theories (Vom Brocke et al. 2021). In addition, the utilization of easily observable digital footprint data to measure the node degree helps to provide beneficial insights into the dynamics of social capital—for instance, how individuals accumulate these valuable assets (e.g., Weiler et al. 2022). As recently corroborated (e.g., Sanchez-Famoso et al. 2020; Saxton and Guo 2020), such research of the concept's dynamics (i.e., processes) is also very much needed to promote our understanding of business phenomena.

7.3 Issues and Peculiarities Associated with Digital Footprint Data

7.3.1 *Validity Issues*

Given the unique nature of digital footprint data, Howison et al. (2011) identified several validity issues that might arise when researchers exploit this kind of data for social network analysis. In the following paragraph, we discuss the issues relevant to our study and place them in light of our findings. Howison et al. (2011) raised awareness of how a specific IS is used by its members. As previously depicted, the networking behavior is quite heterogeneous between individuals and sometimes even within the same individual, as they exhibit diverse strategies depending on whether the amassing of friends and acquaintances is within the real-world or a digital counterpart. In the latter, individuals can utilize the technical possibilities provided by the IS to establish and maintain friendships. For instance, recent experimental evidence suggests that the formation of a digital tie between two members of the IS becomes more likely when information about shared similarities (e.g., having the same hometown) is available (Sun and Taylor 2020) (see also the online appendix for an elaboration on the uniqueness of SMN in social capital processes). Moreover, everyone ascribes a different meaning to the concept of “friend”. For instance, some SMN users might use the low-threshold access to information about other users buzzing around the platform to befriend many people just to artificially boost their network size to stress their social status, while others hand-pick only a few selected friends to signal a feeling of exclusive connectedness (Kane et al. 2014). Thus, researchers should be aware of these actions, as it would be shortsighted to assume that all individuals act uniformly on the platform. Thus, as highlighted by our findings, the subgroup analysis using interaction terms is expedient in gauging social reality. It is also important to visualize that FB is usually not the only IS that participants use in their daily routines, where they leave digital footprints. Individuals usually switch between several channels to communicate with each other, especially if it is a strong tie (Haythornthwaite 2002). Nonetheless, we demonstrated that digital footprint data harvested from FB are sufficient to provide a suitable approximation of the real-world social network size of individuals categorized into certain subgroups. However, researchers should not take this fact at face value and

transfer the evidence unquestioned to other SMNs, as each IS represents a unique ecosystem and repository of digital ties. Compared with specialized SMNs such as LinkedIn or XING, which promote a certain type of digitized relationship (i.e., work-related contacts), FB has no such explicit focus. This implies that on FB, users can potentially befriend, within its system boundaries, people from all walks of life (i.e., family members, close and distant friends, acquaintances, and work colleagues), given that they are also registered members of that specific platform.

Another relevant issue raised by Howison et al. (2011) pertains to the reliability of harvested digital footprint data. While over the years, a platform and its features constantly change, which might affect how the IS is used, the basic mechanism of how users form a digital FB tie (i.e., user A sends a friend request, which user B ideally accepts) and the result of this process (i.e., the increase of the network size of the involved actors by one contact) remains nonetheless the same. Thus, the ways the platform records and displays the user's network size have not changed. We can also reliably assume that this measure correctly reflects the network size of a user within the boundaries of the unique IS, as the platform is a repository that captures all the ties (i.e., strong, weak, and dormant) a user has accumulated since registering. Of course, digital defriending processes are also reflected in that number. Moreover, we also emphasize that given the straightforward operationalization of the structural social capital dimension and its focus on the ego-centric network, researchers have less need to make decisions, such as how to define the tie strength (i.e., relational dimension), which would arise, for example, when studying dyads. In other words, “[p]erhaps because they are relatively familiar objects and more or less fixed over time, the conceptual definition of nodes seems to create fewer problems than the conceptual definition of links [...]” (Howison et al. 2011, p. 776).

Furthermore, we would like to acknowledge that despite differences in data sources and collection, the conceptual foundation of the structural dimension of social capital operationalized with the network-based measure of node degree is unequivocally the same regardless of studying real-world or digital phenomena: “[...] [A] direct connection between the actors x and y can exist only once” (Landherr et al. 2010, p. 377), and the more direct ties people have (i.e., the larger the social network), the more resources they can potentially access (Borgatti et al. 1998)¹². However, we recall the conceptual difference between the node degree harvested from digital footprint data (i.e., SMN-SC) and the counterpart collected through a self-reported contact diary (i.e., O-SC) (see Fig. 1). While the former is comprised of active strong(er) and weak(er) ties and dormant ties, the latter can inherently capture only actively maintained strong(er) and weak(er) ties. Researchers harnessing SMN-SC in their studies should keep this fact in mind and portend to this in their paper.

¹² In that context, we would like to note that it is not unusual that the SMN platform providers implement technical constraints that limit the maximum number of contacts that users can accumulate in their friends list. For instance, Facebook sets the upper limit to 5,000 friends (Kane et al. 2014) and allows the sending of 1,000 contact requests. In the real-world, the numbers of friends and acquaintances are theoretically indefinite. However, constraints, which are cognitive and temporal, exist. These restrictions naturally influence the number of active ties an individual can actually maintain.

7.3.2 Peculiarities

With regard to the potential for digital footprint data to measure the structural dimension of social capital, we also need to consider peculiarities associated with digital SMN data. Sociodemographic profiles and Internet skills affect the likelihood that an individual will use a specific SMN (Blank and Lutz 2017), so potential sources of social capital are inevitably out of reach, if those entities do not participate in SMNs at all. Likewise, it is only possible to retrieve SMN-SC from those people who interact on the specific platform. We do not consider behaviors recorded on another SMN; using another platform to collect digital footprints could result in a different network structure with unique characteristics. Thus, scholars aiming to achieve representative findings must still rely on traditional social science tools. In addition, FB, its features, and its algorithms constantly change, and its popularity could wane, as other SMNs continue to grow. These alterations might lead to shifting profiles of site members (Blank and Lutz 2017), as well as to changes in their behavior (Sundararajan et al. 2013; Kane et al. 2014). If FB users stop using it, perhaps due to envy (Krasnova et al. 2015), but persist in their site membership, reflecting the high transactional costs, the amount of their SMN-SC stagnates, but their amount of O-SC is still subject to continuous change. Moreover, empirical evidence shows that individuals, who maintain larger numbers of digital SMN contacts, were less likely to quit using the platform (Lórinčz et al. 2019). Thus, researchers must understand the data-generating process for the specific SMN platform (Jungherr 2018); changes to it affect which digital footprint data are available (Sundararajan et al. 2013). It also introduces the need to conduct frequent, follow-up indicator validation studies to ensure the reliability of the findings, as without (further) empirical evidence about the convergent validity of indicators, it is not possible to rule out face validity (Jungherr 2018; see also Appel et al. 2014).

One of the biggest influences on the viability of applications of digital footprint data is the access to the data. The platform-owning company usually controls access to the data, as well as the type and amount of harvestable data (Boyd and Crawford 2012; Jungherr 2018). Access to proprietary data also introduces a digital division between researchers (Boyd and Crawford 2012): those granted access to the data by the company or with the technical skills to retrieve it on their own, often via APIs, and those who lack such privileges (Boyd and Crawford 2012; Ruths and Pfeffer 2014). Although the latter could manually retrieve the number of friends by visiting each participant's FB profile page, privacy settings could restrict access to this information (Lankton et al. 2017). Such privacy management choices also vary with sociodemographic factors and could introduce a further bias in the sample variation (see also Hofstra et al. 2017). Moreover, researchers often get only access to a subset of the total amount of available data and are rarely familiar with the underlying selection process (Ruths and Pfeffer 2014), making it difficult to assess the data quality (Boyd and Crawford 2012). These points should sensitize scholars to the methodological challenges with digitalized data. For instance, considering our finding that sociodemographic characteristics are crucial to convergent validity, researchers interested in extracting this information from FB also must realize that some users might fake their profile details to protect their privacy. They must think

carefully about the underlying data and potential biases introduced by the skewed self-selection or platform characteristics, then attempt to mitigate them to arrive at valuable conclusions (Ruths and Pfeffer 2014).

7.4 Limitations and Further Research

A big strength of our study is the use of two unique data sets, which consist of a large number of contact diaries and corresponding digital footprint data of the same participants. Some limitations also deserve acknowledgment. First, the demands to participate in our study were high, in that participants had to provide private FB data and engage in burdensome, time-consuming completion of contact diaries (Fu 2007). To limit the nonresponse rate, we asked participants to keep diaries for one week. Future studies might extend this period to capture more network members (Yen et al. 2016), in that contacts strongly depend on opportunities (e.g., Mollenhorst et al. 2014), seasonal events, holidays, or even weather conditions. Yet, researchers have to keep in mind that extending the period of recordable days might have negative effects on the number of individuals that will finally participate (i.e., the sample size) (Fu 2007; Dávid et al. 2016). In general, Ohly et al. (2010) suggests that researchers considering conducting a diary study should recruit a minimum of 100 individuals to achieve unbiased results. Besides the research presented here, however, only a few prior attempts implementing contact diaries to study egocentric networks actually reach this highly recommended threshold, and they make up the only studies comparable to ours in terms of the length of diary keeping (e.g., Dávid et al. 2016). Most other attempts have sample sizes below the minimum number of 100 recruits (e.g., Fu et al. 2013) because of the previously mentioned tedious and demanding nature of keeping diaries.

Second, some shortcomings are inherent to the diary approach and might have influenced the number of recorded contacts. Participants could (un)intentionally fail to disclose some of their contacts, and the participants themselves determined whether each individual contact was meaningful. This individual subjectivity might lead to over- or underestimates of the amount of O-SC. Nevertheless, we consider this criterion more suitable for capturing active personal networks than other contact definitions, such as those that exclude nonphysical contacts¹³. With our instructions though, some participants recorded occupation groups, such as mail carriers. Thus, we carefully checked our data for such entries, as well as for duplicates, to ensure the robustness of the diary data; we omitted these misrepresented contacts from our data analysis. In addition, participants were more likely not to report brief contacts (Mastrandrea et al. 2015). Because we used a web diary that was easily accessible through any type of devices, participants could record their contacts in real time, which should mitigate recall issues. Furthermore, even with these limitations, contact

¹³ Moreover, it is important to keep in mind that narrowing the instructions on when to record a contact also influences the breadth and depth of the resulting social network. For instance, an instruction that favors the recording of strong ties might make it difficult to capture the relational dimension of social capital adequately, as this dimension should resemble the social ties of all kinds of strengths.

diaries offer the most reliable approach to measure active egocentric networks by concurrently providing detailed information about each social contact (Fu 2007).

Third, the generalizability of our findings might be limited. We rely on a non-representative sample, and we included only the most popular SMN, FB. To improve external validity, we recruited through not just our social network but also in various places and times, which increased sample variation. Follow-up studies also might adopt different sampling methods to recruit participants, such as FB itself or Amazon Mechanical Turk. In addition, continued research could investigate the potential of other sources of digital footprint data, like LinkedIn, for inferring O-SC. By consulting multiple platforms, they could reveal more of each participant's digitized social capital. To increase the generalizability of our findings, studies also should extend the geographical focus beyond Germany; other countries might feature different FB adoption rates and unique user bases.

Fourth, Kossinets and Watts (2006) showed that longitudinal digital footprint data (i.e., e-mail exchanges) have proven to be beneficial in understanding the evolution of social networks. While the underlying digital footprint data of their study do not exploit SMN data but e-mail exchanges, their results, nonetheless, also suggest some important aspects for our study. In particular, their findings stress the dynamic nature of networks and that over time, an individual does not always occupy the same position within these social structures (i.e., suggesting the instability of individual network properties). However, these individual changes seem to level each other out so that, on average, the network properties exhibit stability. Regarding the purpose of our study, this indicates that exploiting cross-sectional digital footprint data is expedient if they are the most recent version of it because then they can adequately capture the prevailing individual differences. Moreover, given that Kossinets and Watts (2006) used data obtained from e-mail exchanges instead of SMN data, their utilized data source reflects a different social reality, as this collaboration technology (i.e., e-mail) “[...] do not allow users to establish profiles or lists of connections that can be viewed or traversed by others” (Kane et al. 2014, p. 279). Thus, future studies should be aimed at generating further knowledge on that topic and utilize different snapshots of the SMN node degree (e.g., before, during, and after the collection of the contact diary data) and relate them to the real-world counterpart. This would help assess how potentially changing digital network structures affect the convergent validity of the measure.

Fifth, in line with other researchers (Appel et al. 2014), who also vetted the convergent validity of different social capital measures, we regard a correlation coefficient of 0.31 or higher as sufficiently strong to suggest a reliable proxy measure. However, we acknowledge that some scholars might endorse a more conservative approach and consider only stronger correlation coefficients (>0.7) as expedient to establish convergent validity. If we applied such a strict cutoff value, it would imply that SMN-SC only works well for the subgroup of young females as an appropriate proxy measure.

On a related note, individual behavior does not follow deterministic rules nor is it homogenous per definition, rather it resolves in heterogenous actions. Among other characteristics, we can regard one's age and gender—which shape how individuals act in the digital sphere or real-world, as pivotal—as highlighted in the

present study by the essential role of the identified three-way interaction. Yet, critics may demur that our main effect is only significant because of the inclusion of the control variables and interactions and therefore might be an artifact caused by statistical suppression effects. By contrast, we can, however, contend that if a finding is based on statistical suppression effects, it would be fragile to the specific model specification (Lenz and Sahn 2021). Given that we estimated different specified regressions models, which all lead to robust results, we are confident that we can invalidate such claims. Moreover, the inclusion of rather unusual control variables, such as posttreatment variables, is also suspected to be a focal vehicle for introducing statistical suppression effects (Lenz and Sahn 2021). Explicitly, the controls we exploited, in particular, our essential sociodemographic variables—age and gender—are no contender for falling into this category of being regarded as unusual control variables. Moreover, excluding such important characteristics of an individual that are known to shape its behavior might rather introduce omitted variable bias. Nonetheless, further studies are warranted to validate our findings or report different insights. Either way, both outcomes would strengthen the epistemological value of exploiting a measure harvested from digital footprint data as a proxy for its self-reported counterpart.

Finally, although the notions of the terms we defined (i.e., SMN-SC and O-SC) might suggest otherwise, as highlighted throughout the present manuscript, we focused exclusively on measuring the structural dimension of the concept of social capital harnessing digital footprint data. We operationalized this scaffolding dimension of the concept using the node degree, which captures the size of a person's maintained social network. As mentioned in the Section "Theoretical Background", social capital consists of several dimensions, and it should be the guiding light—if implementable—when leveraging the concept in research to include all those dimensions in one's study. However, in some cases it might only be possible to utilize the foundational dimension of structural social capital. Against that backdrop, scholars should keep in mind that the node degree and the sheer network size it represents provide a rather broad measure of the concept of social capital. Thus, the obtainable insights should be more fine-grained if the concepts other two dimensions are taken into account as well. The cognitive dimension of social capital would be difficult to measure with digital footprint data, because self-reported data are needed to capture this dimension (e.g., Tsai and Ghoshal 1998). Nevertheless, we could model the relational dimension with FB interaction data, such as the number of exchanged messages (e.g., Arnaboldi et al. 2013). Compared with the centrality metric node degree, which is often publicly visible or retrievable with basic permissions, FB interaction data require extended permissions. Moreover, as previously depicted, studying the tie-strength also comes with certain decisions that researchers must face, such as how to define a link between two individuals. Still, future research should also take the quality of the social relationships into account and not just the existence of ties, because tie strength is associated with different benefits (e.g., Levin and Cross 2004). Evidence based on self-reported data suggests that digitized and real-world ties do not significantly differ in terms of tie strength, in support of the potential benefits of digital footprint data (Bisbee and Larson 2017). Further research is necessary to reaffirm this finding with actual digital footprint data. In addition,

in the identified cases, both metrics of structural social capital are informative for each other, we have reason to believe that their effect on an outcome measure would be similar, as noted in the introduction. We call on further research to consider this relationship explicitly, by comparing the ability of SMN-SC and O-SC to influence outcome measures such as career success or other economic outcomes (e.g., Hinz and Spann 2008); such findings could strengthen our results too.

Despite these shortcomings, our study offers an important first assessment of the convergent validity of SMN-SC in measuring O-SC, or, in other words, the potential of the node degree extracted from the digital footprint data in facilitating the measurement of the structural dimension of social capital. We hope it encourages researchers to conduct further proxy indicator validation studies, as the verification of the potential of digital footprint data, in terms of measuring the structural dimension of social capital and its other dimensions, is critical, especially as human behavior continues to materialize itself in digital spheres.

Supplementary Information The online version of this article (<https://doi.org/10.1007/s41471-024-00180-8>) contains supplementary material, which is available to authorized users.

Funding The work was funded by the DFG under the Graduiertenkolleg 2050 (Privacy and Trust for Mobile Users) funding.

Open Access funding enabled and organized by Schmalenbach-Gesellschaft and German Academic Association for Business Research.

Declarations

Conflict of interest M. Weiler, N. Jansen and O. Hinz declare that they have no competing interests.

Ethical standards For this article no studies with human participants or animals were performed by any of the authors. All studies mentioned were in accordance with the ethical standards indicated in each case. We hereby declare that we complied with the Ethical Standards as outlined on the SBUR webpage.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adler, P.S., and S.-W. Kwon. 2002. Social capital: prospects for a new concept. *Academy of Management Review* 27(1):17–40.
- Alhabash, S., and M. Ma. 2017. A tale of four platforms: motivations and uses of Facebook, twitter, Instagram, and Snapchat among college students? *Social Media+ Society* <https://doi.org/10.1177/2056305117691544>.
- Appel, L., P. Dadlani, M. Dwyer, K. Hampton, V. Kitzie, Z.A. Matni, P. Moore, and R. Teodoro. 2014. Testing the validity of social capital measures in the study of information and communication technologies. *Information, Communication & Society* 17(4):398–416.

- Arnaboldi, V., A. Guazzini, and A. Passarella. 2013. Egocentric online social networks: analysis of key features and prediction of tie strength in Facebook. *Computer Communications* 36(10):1130–1144.
- Bisbee, J., and J.M. Larson. 2017. Testing social science network theories with online network data: an evaluation of external validity. *American Political Science Review* 111(3):502–521.
- Blank, G., and C. Lutz. 2017. Representativeness of social media in Great Britain: investigating Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram. *American Behavioral Scientist* 61(7):741–756.
- Boeschoten, L., J. Ausloos, J. Moeller, T. Araujo, and D.L. Oberski. 2020. *Digital trace data collection through data donation*. arXiv, Vol. 201109851 arXiv preprint.
- Böger, A., M. Wetzel, and O. Huxhold. 2017. Allein unter vielen oder zusammen ausgeschlossen: Einsamkeit und wahrgenommene soziale Exklusion in der zweiten Lebenshälfte. In *Altern im Wandel : zwei Jahrzehnte Deutscher Alterssurvey (DEAS)*, ed. K. Mahne, J.K. Wolff, J. Simonson, and C. Tesch-Römer, 273–285. Wiesbaden: Springer.
- Bolino, M.C., W.H. Turnley, and M.B. James. 2002. Citizenship behavior and the creation of social capital in organizations. *The Academy of Management Review* 27(4):505–522.
- Borgatti, S.P., C. Jones, and M.G. Everett. 1998. Network measures of social capital. *Connections* 21(2):27–36.
- Bourdieu, P. 1986. The forms of capital. In *Handbook of theory and research for the sociology of education*, Vol. 1, ed. J. Richardson, 241–258. Westport: Greenwood.
- Boyd, D., and K. Crawford. 2012. Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15(5):662–679.
- Brandtzaeg, P.B. 2012. Social networking sites: their users and social implications—A longitudinal study. *Journal of Computer-Mediated Communication* 17(4):467–488.
- Brashears, M.E., E. Hoagland, and E. Quintane. 2016. Sex and network recall accuracy. *Social Networks* 44:74–84.
- Carlson, K.D., and A.O. Herdman. 2012. Understanding the impact of convergent validity on research results. *Organizational Research Methods* 15(1):17–32.
- Chen, G.M. 2014. Revisiting the social enhancement hypothesis: extroversion indirectly predicts number of Facebook friends operating through Facebook usage. *Computers in Human Behavior* 39:263–269.
- Coleman, J.S. 1988. Social capital in the creation of human capital. *American Journal of Sociology* 94:S95–S120.
- Cornwell, B., E.O. Laumann, and L.P. Schumm. 2008. The social connectedness of older adults: a national profile. *American Sociological Review* 73(2):185–203.
- Correa, T. 2016. Digital skills and social media use: how Internet skills are related to different types of Facebook use among 'digital natives. *Information, Communication & Society* 19(8):1095–1107.
- Couper, M.P. 2017. New developments in survey data collection. *Annual Review of Sociology* 43(1):121–145.
- Cummings, J., and A.R. Dennis. 2018. Virtual first impressions matter: the effect of enterprise social networking sites on impression formation in virtual teams. *MIS Quarterly* 42(3):697–717.
- Dávid, B., E. Huszti, I. Barna, and Fu Y-c. 2016. Egocentric contact networks in comparison: Taiwan and Hungary. *Social Networks* 44:253–265.
- DiPrete, T.A., A. Gelman, T. McCormick, J. Teitler, and T. Zheng. 2011. Segregation in social networks based on acquaintanceship and trust. *American Journal of Sociology* 116(4):1234–1283.
- Dissanayake, I., J. Zhang, and B. Gu. 2015. Task division for team success in crowdsourcing contests: resource allocation and alignment effects. *Journal of Management Information Systems* 32(2):8–39.
- Dunbar, R.I. 2016. Do online social media cut through the constraints that limit the size of offline social networks? *Open Science* 3(1):150292.
- Dunbar, R.I. 2018. The anatomy of friendship. *Trends in Cognitive Sciences* 22(1):32–51.
- Dunbar, R.I., V. Arnaboldi, M. Conti, and A. Passarella. 2015. The structure of online social networks mirrors those in the offline world. *Social Networks* 43:39–47.
- Eagle, N., A.S. Pentland, and D. Lazer. 2009. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences of the United States of America* 106(36):15274–15278.
- Eagly, A.H., W. Wood, and A.B. Diekmann. 2000. Social role theory of sex differences and similarities: a current appraisal. In *The developmental social psychology of gender*, ed. T. Eckes, H.M. Trautner, 123–174. Mahwah: Erlbaum.
- Eklund, L. 2015. Bridging the online/offline divide: the example of digital gaming. *Computers in Human Behavior* 53:527–535.
- Ellison, N., and J. Vitak. 2015. Social network site affordances and their relationship to social capital processes. In *The handbook of the psychology of communication technology*, ed. S.S. Sundar, 205–237. Hoboken: Wiley-Blackwell.

- Ellison, N.B., C. Steinfield, and C. Lampe. 2007. The benefits of Facebook “friends:” social capital and college students’ use of online social network sites. *Journal of Computer-Mediated Communication* 12(4):1143–1168.
- Ellison, N.B., J. Vitak, R. Gray, and C. Lampe. 2014. Cultivating social resources on social network sites: Facebook relationship maintenance behaviors and their role in social capital processes. *Journal of Computer-Mediated Communication* 19(4):855–870.
- Engbers, T.A., M.F. Thompson, and T.F. Slaper. 2017. Theory and measurement in social capital research. *Social Indicators Research* 132(2):537–558.
- Flap, H., and B. Völker. 2001. Goal specific social capital and job satisfaction effects of different types of networks on instrumental and social aspects of work. *Social Networks* 23:297–320.
- Florin, J., M. Lubatkin, and W. Schulze. 2003. A social capital model of high-growth ventures. *Academy of Management Journal* 46(3):374–384. <https://doi.org/10.5465/30040630>.
- Fu, Y. 2007. Contact diaries: building archives of actual and comprehensive personal networks. *Field Methods* 19(2):194–217.
- Fu, Y.-c., H.-C. Ho, and H.M. Chen. 2013. Weak ties and contact initiation in everyday life: Exploring contextual variations from contact diaries. *Soc Network* 35(3):279–287.
- Gaito, S., G.P. Rossi, and M. Zignani. 2012. Facencounter: bridging the gap between offline and online social networks. In *Paper presented at the Proceedings of the 2012 Eighth International Conference on Signal Image Technology and Internet Based Systems* <https://doi.org/10.1109/SITIS.2012.116>.
- Gil de Zúñiga, H., N. Jung, and S. Valenzuela. 2012. Social media use for news and individuals’ social capital, civic engagement and political participation. *Journal of Computer-Mediated Communication* 17(3):319–336.
- Glüer, M., and A. Lohaus. 2016. Participation in social network sites: associations with the quality of offline and online friendships in German preadolescents and adolescents. *Cyberpsychology* 10(2). <https://doi.org/10.5817/CP2016-2-2>.
- Golder, S.A., and M.W. Macy. 2014. Digital footprints: opportunities and challenges for online social research. *Annual Review of Sociology* 40(1):129–152.
- Greenwood, S., A. Perrin, and M. Duggan. 2016. Social media update 2016. Pew research center. <http://www.pewinternet.org/2016/11/11/social-media-update-2016/>. Accessed 15 Dec 2017.
- Hahn, E., J. Gottschling, and F.M. Spinath. 2012. Short measurements of personality-validity and reliability of the GSOEP Big Five Inventory (BFI-S). *Journal of Research in Personality* 46(3):355–359.
- Haythornthwaite, C. 2002. Strong, weak, and latent ties and the impact of new media. *The Information Society* 18(5):385–401. <https://doi.org/10.1080/01972240290108195>.
- Herz, A., and S. Petermann. 2017. Beyond interviewer effects in the standardized measurement of ego-centric networks. *Social Networks* 50:70–82.
- Hill, R.A., and R.I. Dunbar. 2003. Social network size in humans. *Human Nature* 14(1):53–72.
- Hinz, O., and M. Spann. 2008. The impact of information diffusion on bidding behavior in secret reserve price auctions. *Information Systems Research* 19(3):351–368.
- Hinz, O., B. Skiera, C. Barrot, and J.U. Becker. 2011. Seeding strategies for viral marketing: an empirical comparison. *Journal of Marketing* 75(6):55–71.
- Hinz, O., M. Spann, and I.-H. Hann. 2015. Research note—Can’t buy me love... or can I? Social capital attainment through conspicuous consumption in virtual environments. *Information Systems Research* 26(4):859–870.
- Hofstra, B., R. Corten, F. van Tubergen, and N.B. Ellison. 2017. Sources of segregation in social networks: a novel approach using Facebook. *American Sociological Review* 82(3):625–656.
- Howison, J., A. Wiggins, and K. Crowston. 2011. Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems* 12(12):767–797.
- Jansen, N., and O. Hinz. 2022. Inferring opinion leadership from digital footprints. *Journal of Business Research* 139:1123–1137. <https://doi.org/10.1016/j.jbusres.2021.10.032>.
- Johnson, R., B. Kovács, and A. Vicsek. 2012. A comparison of email networks and off-line social networks: a study of a medium-sized bank. *Social Networks* 34(4):462–469.
- Joinson, A.N. 2008. Looking at, looking up or keeping up with people?: motives and use of Facebook. In *Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy*.
- Junco, R. 2013. Inequalities in Facebook use. *Computers in Human Behavior* 29(6):2328–2336.
- Jungherr, A. 2018. Normalizing digital trace data. In *Digital discussions: how big data informs political communication*, ed. N.J. Stroud, S. McGregor, 9–35. New York: Routledge.

- Jungherr, A., H. Schoen, O. Posegga, and P. Jürgens. 2017. Digital trace data in the study of public opinion: an indicator of attention toward politics rather than political support. *Social Science Computer Review* 35(3):336–356.
- Kalmijn, M. 2012. Longitudinal analyses of the effects of age, marriage, and parenthood on social contacts and support. *Advances in Life Course Research* 17(4):177–190.
- Kane, G.C., M. Alavi, G. Labianca, and S.P. Borgatti. 2014. What's different about social media networks? A framework and research agenda. *MIS Quarterly* 38(1):275–304.
- Katz, E., and P.F. Lazarsfeld. 1955. *Personal influence: the part played by people in the flow of mass communications*. Free Press.
- Kibanov, M., M. Atzmueller, J. Illig, C. Scholz, A. Barrat, C. Cattuto, and G. Stumme. 2015. Is web content a good proxy for real-life interaction? A case study considering online and offline interactions of computer scientists. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 25–28 Aug. 2015, 697–704.
- Kosinski, M., D. Stillwell, and T. Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America* 110(15):5802–5805.
- Kosinski, M., S.C. Matz, S.D. Gosling, V. Popov, and D. Stillwell. 2015. Facebook as a research tool for the social sciences: opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist* 70(6):543–556.
- Kossinets, G., and D.J. Watts. 2006. Empirical analysis of an evolving social network. *Science* 311(5757): 88–90. <https://doi.org/10.1126/science.1116869>.
- Krämer, N.C., V. Sauer, and N. Ellison. 2021. The strength of weak ties revisited: further evidence of the role of strong ties in the provision of online social support. *Social Media + Society* 7(2):20563051211024958. <https://doi.org/10.1177/20563051211024958>.
- Krasnova, H., T. Widjaja, P. Buxmann, H. Wenninger, and I. Benbasat. 2015. Research note—Why following friends can hurt you: an exploratory investigation of the effects of envy on social networking sites among college-age users. *Information Systems Research* 26(3):585–605.
- Krasnova, H., N.F. Veltri, N. Eling, and P. Buxmann. 2017. Why men and women continue to use social networking sites: The role of gender differences. *The Journal of Strategic Information Systems* 26(4):261–284.
- Landherr, A., B. Friedl, and J. Heidemann. 2010. A critical review of centrality measures in social networks. *Business & Information Systems Engineering* 2(6):371–385. <https://doi.org/10.1007/s12599-010-0127-3>.
- Lankton, N.K., D.H. McKnight, and J.F. Tripp. 2017. Facebook privacy management strategies: a cluster analysis of user privacy behaviors. *Computers in Human Behavior* 76:149–163.
- Lenz, G.S., and A. Sahn. 2021. Achieving statistical significance with control variables and without transparency. *Political Analysis* 29(3):356–369.
- Levin, D.Z., and R. Cross. 2004. The strength of weak ties you can trust: the mediating role of trust in effective knowledge transfer. *Management Science* 50(11):1477–1490.
- Levin, D.Z., J. Walter, and J.K. Murnighan. 2011. The power of reconnection—How dormant ties can surprise you. *MIT Sloan Management Review* 52(3):45–50.
- Li, Y., X. Wang, L. Huang, and X. Bai. 2013. How does entrepreneurs' social capital hinder new business development? A relational embeddedness perspective. *Journal of Business Research* 66(12):2418–2424. <https://doi.org/10.1016/j.jbusres.2013.05.029>.
- Lin, N. 1999. Building a network theory of social capital. *Connections* 22(1):28–51.
- Liu, D., and W.K. Campbell. 2017. The big five personality traits, big two metatraits and social media: a meta-analysis. *Journal of Research in Personality* 70:229–240.
- Lizardo, O., and S. Skiles. 2012. Reconceptualizing and theorizing 'omnivorousness': genetic and relational mechanisms. *Sociological Theory* 30(4):263–282.
- Lőrincz, L., and B. Németh. 2022. How social capital is related to migration between communities? *European Journal of Population* 38(5):1119–1143.
- Lőrincz, L., J. Koltai, A.F. Győr, and K. Takács. 2019. Collapse of an online social network: burning social capital to create it? *Social Networks* 57:43–53.
- Loury, G. 1992. The economics of discrimination: getting to the core of the problem. *Harvard Journal for African American Public Policy* 1(1):91–110.
- Lüdeke, H., and H. Allinger. 2017. Zeig mir deine Freunde und ich sag dir, wer du bist – Ein empirischer Test zur Berücksichtigung der Akteursheterogenität in der Sozialkapitalforschung. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung* 69(1):1–40. <https://doi.org/10.1007/s41471-016-0020-y>.

- Marsden, P.V. 2011. Survey methods for network data. In *The SAGE handbook of social network analysis*, Vol. 25, ed. J. Scott, P.J. Carrington, 370–388. London: SAGE.
- Mastrandrea, R., J. Fournet, and A. Barrat. 2015. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLOS ONE* 10(9):e136497. <https://doi.org/10.1371/journal.pone.0136497>.
- McAndrew, F.T., and H.S. Jeong. 2012. Who does what on Facebook? Age, sex, and relationship status as predictors of Facebook use. *Computers in Human Behavior* 28(6):2359–2365.
- McDonald, S., and C.A. Mair. 2010. Social capital across the life course: age and gendered patterns of network resources. *Sociological Forum* 25(2):335–359.
- Mollenhorst, G., B. Volker, and H. Flap. 2014. Changes in personal relationships: how social contexts affect the emergence and discontinuation of relationships. *Social Networks* 37:65–80.
- Munch, A., J.M. McPherson, and L. Smith-Lovin. 1997. Gender, children, and social contact: the effects of childrearing for men and women. *American Sociological Review* 62(4):509–520.
- Muscianell, N.L., and R.E. Guadagno. 2012. Make new friends or keep the old: gender and personality differences in social networking use. *Computers in Human Behavior* 28(1):107–112.
- Nahapiet, J., and S. Ghoshal. 1998. Social capital, intellectual capital, and the organizational advantage. *The Academy of Management Review* 23(2):242–266.
- Norbutas, L., and R. Corten. 2018. Network structure and economic prosperity in municipalities: a large-scale test of social capital theory using social media data. *Social Networks* 52:120–134.
- OECD. 2015. *ISCED 2011 operational manual—Guidelines for classifying national education programmes and related qualifications*. Paris: OECD.
- Oh, H.J., C. Lauckner, J. Boehmer, R. Fewins-Bliss, and K. Li. 2013. Facebooking for health: an examination into the solicitation and effects of health-related social support on social networking sites. *Computers in Human Behavior* 29(5):2072–2080.
- Ohly, S., S. Sonnentag, C. Niessen, and D. Zapf. 2010. Diary studies in organizational research: an introduction and some practical recommendations. *Journal of Personnel Psychology* 9(2):79–93.
- Payne, G.T., C.B. Moore, S.E. Griffis, and C.W. Autry. 2011. Multilevel challenges and opportunities in social capital research. *Journal of Management* 37(2):491–520.
- Quan-Haase, A., G.Y. Mo, and B. Wellman. 2017. Connected seniors: how older adults in East York exchange social support online and offline. *Information, Communication & Society* 20(7):967–983.
- Roberts, S.G.B., R. Wilson, P. Fedurek, and R.I.M. Dunbar. 2008. Individual differences and personal social network size and structure. *Personality and Individual Differences* 44(4):954–964.
- Roberts, S.G.B., R.I.M. Dunbar, T.V. Pollet, and T. Kuppens. 2009. Exploring variation in active network size: constraints and ego characteristics. *Social Networks* 31(2):138–146.
- Ruths, D., and J. Pfeffer. 2014. Social media for large studies of behavior. *Science* 346(6213):1063–1064.
- Sanchez-Famoso, V., A. Maseda, T. Iturralde, S.M. Danes, and G. Aparicio. 2020. The potential of internal social capital in organizations: an assessment of past research and suggestions for the future. *Journal of Small Business Management* 58(1):32–72.
- Saxton, G.D., and C. Guo. 2020. Social media capital: conceptualizing the nature, acquisition, and expenditure of social media-based organizational resources. *International Journal of Accounting Information Systems* <https://doi.org/10.1016/j.accinf.2019.100443>.
- Schober, M.F., J. Pasek, L. Guggenheim, C. Lampe, and F.G. Conrad. 2016. Social media analyses for social measurement. *Public Opinion Quarterly* 80(1):180–211.
- Selden, M., and A.S. Goodie. 2018. Review of the effects of five factor model personality traits on network structures and perceptions of structure. *Social Networks* 52:81–99.
- Skiera, B., O. Hinze, and M. Spann. 2015. Social media and academic performance: does the intensity of Facebook activity relate to good grades? *Schmalenbach Business Review* 67(1):54–72.
- Socievole, A., F. De Rango, and A. Caputo. 2016. Opportunistic mobile social networks: from mobility and Facebook friendships to structural analysis of user social behavior. *Computer Communications* 87:1–18.
- Statistisches Bundesamt. 2016. *Statistik und Wissenschaft. Demographische Standards, Ausgabe 2016*. http://www.destatis.de/GPStatistik/servlets/MCRFileNodeServlet/DEMonografie_derivate_00001549/Band17_DemographischeStandards1030817169004.pdf.
- von Stetten, A., D. Beimborn, and T. Weitzel. 2012. Analyzing and managing the impact of cultural behavior patterns on social capital in multinational IT project teams. *Business & Information Systems Engineering* 4(3):137–151.
- Stier, S., J. Breuer, P. Siegers, and K. Thorson. 2020. Integrating survey data and digital trace data: key issues in developing an emerging field. *Social Science Computer Review* 38(5):503–516.

- Stopczynski, A., V. Sekara, P. Sapiezynski, A. Cuttone, M.M. Madsen, J.E. Larsen, and S. Lehmann. 2014. Measuring large-scale social networks with high resolution. *PLOS ONE* 9(4):e95978. <https://doi.org/10.1371/journal.pone.0095978>.
- Sun, T., and S.J. Taylor. 2020. Displaying things in common to encourage friendship formation: a large randomized field experiment. *Quantitative Marketing and Economics* 18:237–271. <https://doi.org/10.1007/s11129-020-09224-9>.
- Sundararajan, A., F. Provost, G. Oestreicher-Singer, and S. Aral. 2013. Research commentary—Information in digital, economic, and social networks. *Information Systems Research* 24(4):883–905.
- Tabachnick, B.G., and L.S. Fidell. 2007. *Using multivariate statistics*. Boston, New York, Paris: Pearson Education.
- Tsai, W., and S. Ghoshal. 1998. Social capital and value creation: the role of intrafirm networks. *Academy of Management Journal* 41(4):464–476.
- Umberson, D., M.D. Chen, J.S. House, K. Hopkins, and E. Slaten. 1996. The effect of social relationships on psychological well-being: are men and women really so different? *American Sociological Review* 61(5):837–857.
- Van Dijk, J.A. 2013. A theory of the digital divide. In *The digital divide: the Internet and social inequality in international perspective*, ed. M. Ragnedda, G.W. Muschert, 29–51. Abingdon: Routledge.
- Van Der Gaag, M., and T.A.B. Snijders. 2005. The resource generator: social capital quantification with concrete items. *Social Networks* 27(1):1–29.
- Van Deursen, A.J.A.M., E.J. Helsper, and R. Eynon. 2016. Development and validation of the Internet Skills Scale (ISS). *Information, Communication & Society* 19(6):804–823.
- Van Zalk, M.H.W., N. Van Zalk, M. Kerr, and H. Stattin. 2014. Influences between online-exclusive, conjoint and offline-exclusive friendship networks: the moderating role of shyness. *European Journal of Personality* 28(2):134–146.
- Vom Brocke, J., W. Van der Aalst, T. Grisold, W. Kremser, J. Mendling, B. Pentland, J. Recker, M. Roeglinger, M. Rosemann, and B. Weber. 2021. *Process science: the interdisciplinary study of continuous change*. <https://doi.org/10.2139/ssrn.3916817>.
- Von Der Heide, R., G. Vyas, and I.R. Olson. 2014. The social network-network: size is predicted by brain structure and function in the amygdala and paralimbic regions. *Social Cognitive and Affective Neuroscience* 9(12):1962–1972.
- Weiler, M., S. Stolz, A. Lanz, C. Schlereth, and O. Hinz. 2022. Social capital accumulation through social media networks: evidence from a randomized field experiment and individual-level panel data. *MIS Quarterly* 46(2):771–812. <https://doi.org/10.25300/MISQ/2022/16451>.
- Williams, R. 2012. Using the margins command to estimate and interpret adjusted predictions and marginal effects. *Stata Journal* 12(2):308–331.
- Yen, T.-J., Y. Fu, and J.-S. Hwang. 2016. Alters as species: predicting personal network size from contact diaries. *Social Networks* 45:78–88.
- Zywica, J., and J. Danowski. 2008. The faces of Facebookers: investigating social enhancement and social compensation hypotheses; predicting Facebook™ and offline popularity from sociability and self-esteem, and mapping the meanings of popularity with semantic networks. *Journal of Computer-Mediated Communication* 14(1):1–34.

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