Promises and Pitfalls of Instagram Data Donations

Irene I. van Driel[1], Anastasia Giachanou[2], J. Loes Pouwels[3], Laura Boeschoten[2], Ine Beyens[1], & Patti M. Valkenburg[1]

[1] Amsterdam School of Communication Research, University of Amsterdam. [2] Department of Methodology and Statistics, Utrecht University, [3] Behavioural Science Institute, Radboud University Nijmegen

Author Notes

This study was funded by an NWO Spinoza Prize and a Gravitation grant (NWO Grant 024.001.003; Consortium on Individual Development) awarded to Patti Valkenburg by the Dutch Research Council (NWO). Corresponding author: Irene van Driel, i.i.vandriel@uva.nl
We would like to thank Tim Verbeij, Teun Siebers, and Wieneke Rollman for their contribution to the data collection and/or coding of this study.

Credit Author Statement:

Irene van Driel: Conceptualization, data collection and coding, writing, revising original and subsequent drafts. **Anastasia Giachanou:** Script building and analysis of DDP data, creating graphs, reviewing drafts. **Loes Pouwels**: Conceptualization, methodological advice, revising original draft. **Laura Boeschoten**: Conceptualization, script building for anonymizing DDPs, reviewing final draft. **Ine Beyens**: Conceptualization, reviewing final draft. **Patti Valkenburg**: Conceptualization, writing, revising original and subsequent drafts.

Please cite as: van Driel, I. I., Giachanou, A., Pouwels, J., Boeschoten, L., Beyens, I., & Valkenburg, P. M. (2021, December 9). Promises and Pitfalls of Instagram Data Donations. https://doi.org/10.31219/osf.io/krqb9

Abstract

Studies into the association of social media use with mental health are largely based on measures of time spent on social media. The small and inconsistent results in these studies may be due to a lack of explanatory power of time-based measures. Data Download Packages (DDPs), the archives of social media platforms that each user is allowed to download, provide a new and promising method to collect content-based information about social media use. In this study, we discuss the promises and pitfalls of DDPs based on an exploratory analysis of 110 Instagram DDPs gathered from 102 adolescents. DDPs provide tremendous opportunities to get insight in the frequency, range, and content of social media activities, from browsing to searching and posting. Yet, the method is also complex and laborious and demands numerous procedural and analytical choices and decisions. Moreover, due to several factors unique to social media interactions, automated content analysis may be challenging. We successively discuss the challenges and opportunities of collecting and analyzing DDPs to help future researchers in their consideration of whether and how to use DDPs.

Promises and Pitfalls of Instagram Data Donations

How social media use affects people's mental health is a research question in progress. Thus far, most studies have focused on self-reported time spent on social media platforms as an indicator of social media use. The effects of social media use on mental health reported in these studies have been small and inconsistent (Appel et al., 2020; Coyne et al., 2020; Dienlin & Johannes, 2020; Odgers & Jensen, 2020; Orben, 2020). These mixed findings may be due to the heterogeneous nature of social media effects (Beyens et al., 2020), but may also point to a more general lack of explanatory power of self-reported time-based predictors of mental health (Odgers & Jensen, 2020; Valkenburg et al., 2022). Time-based measures of social media use do not provide insight in what people see or do on social media. After all, people may engage with social media in different ways, and respond to social media content in different ways (Griffioen et al., 2021; Rideout & Fox, 2018).

To address the limitations of self-reported time-based measures of social media use, researchers are calling for new studies that contribute to knowledge on the type of activities social media users engage in (e.g., private messaging and public posting) and the content they share and encounter (e.g., positively versus negatively valenced messages) (Griffioen et al., 2020; Odgers & Jensen, 2020). In response to these calls, a variety of methodologies that assess social media use in a more objective manner have been explored, most of which utilize smartphone devices as measurement tool (Burnell et al., 2021; Lind et al., 2018; Reeves et al., 2021). Although promising and necessary, these objective measurements also have their limitations, such as that they yield social media data that are confined to smartphone use only.

The current study is the first to explore the promises and pitfalls of a new method to capture social media use based on data donations of social media archives: Data Download Packages (DDPs). Since the implementation in Europe of the General Data Protection Regulation in 2018, all platforms that store data of their users are legally mandated to share

these data with their users upon their request (Boeschoten et al., 2020). This paper explores the process of data donations of Instagram, a highly popular platform that offers a wide range of ways to engage with. The first aim of this study is to share how Instagram DDPs can be obtained and the second aim is to share what these DDPs can teach us about Instagram use. While this study focuses on adolescents, the most avid users of Instagram, the experiences we share also largely apply to studies among other age groups.

How Data Donations Could Complement Existing Measures of Social Media Use

The potential of several objective social media use measures has been investigated in recent years (Griffioen et al., 2020). In particular, research in which smartphone use is tracked is a major step forward and has uncovered meaningful discrepancies between self-report and smartphone logs (Johannes et al., 2021; Parry et al., 2021; Verbeij et al., 2021). Unfortunately though, most of these log-based measures are still limited to measuring overall time spent on social media, and, thus, provide little insight in the specific activities and content that is shared and encountered (Valkenburg et al., 2022).

Acquiring objective social media use data that provide further insight in specific activities and content shared and encountered on social media platforms comes with obvious challenges as to how to access this content while safeguarding participants' privacy. Researchers have found innovative solutions to this challenge, for example by collecting content of social media accounts with the help of application programming interfaces (API's) provided by the platforms (Bayer et al., 2018). A main drawback of this method is that it only provides access to the type of data that are selected by the platform and is restricted to publicly available data. This method works for platforms in which profiles are automatically set to public, such as Twitter, but not for platforms such as Facebook or Instagram where users can select their privacy settings (Batrinca & Treleaven, 2015).

Other promising objective approaches are mobile sensing (Lind et al., 2018) and *Screenomics* (Ram et al., 2019). Mobile sensing tools are custom-built applications that collect data through existing smartphone sensors (e.g., GPS, Bluetooth, microphone). Some mobile sensing tools make use of additional sensors that are relevant for measuring social media use. For example, the tool *EARS* adds a custom keyboard that captures all text entered through that keyboard (Lind et al., 2018). The *Screenomics* approach extracts visual as well as textual content by taking automated screenshots of participants' smartphone every few seconds (Ram et al., 2019).

Although both the mobile sensing and the *Screenomics* approach hold great promise to capture moment-to-moment involvement with social media, DDPs come with four additional advantages. First, DDPs provide a full overview of the uses of a platform regardless of whether it was accessed via the phone, tablet, or laptop. Second, DDPs capture all user interactions with the platform from the moment a user created an account until the moment of the download request. Third, because data of platform users are collected automatically by the social media companies, no (smartphone) applications need to be installed and thus researcher bias is limited. Fourth, all information is timestamped and separated into text and media files, categorized per social media activity.

To obtain initial insights in the promises and pitfalls of the DDP method, we focus on Instagram because it affords a wide range of user activities that are also characteristic of other platforms, such as (a) engaging in private exchanges through direct messaging, (b) browsing through profiles and content of others, (c) sharing content publicly by means of posts, and (d) sharing stories, which represent ephemeral content that disappears after 24 hours of posting. Moreover, Instagram DDPs provide important other types of information such as the date that the account was first created, the number of followers, and whether the account is private or public.

To obtain insight in how Instagram data donations can be obtained (Aim 1), we will describe the participant recruitment process and how we instructed participants to download and share their DDPs. To obtain insight in what these DDPs can teach us about Instagram use (Aim 2), we will describe findings from an initial analysis of the DDPs, starting with descriptions of the Instagram accounts (e.g., number of followers) and subsequently of the types and frequencies of the Instagram activities participants engaged in. Finally, we describe the results of a preliminary comparison of a manual and automated coding of the sentiment (the positivity or negativity) of the posts included in the DDPs.

Obtaining Data Donation Packages (Aim 1)

The Instagram data donation study is part of a larger research project on social media use and mental health that ran from 21 November 2019 to 1 July 2020. The data donation was the last component of this project and took place at the end of June and beginning of July 2020. An overview of the timeline of the overarching project is available on the Open Science Framework (https://osf.io/n8v9f/?view_only=e4573d0dbfc745b79c9407f3612e0966).

Participants

Of the 388 participants who started the overall project, 102 participants provided 110 useable Instagram DDPs: 96 participants donated one account, four donated two accounts, and two shared three accounts. The sample of participants who donated their DDPs was comparable to the total sample in terms of age (*M* age = 14.04 vs *M* age total sample = 14.11). But the DDP sample consisted of more girls (67% vs 54% of total sample), more participants who followed a higher educational track at school (34% vs 26% of total sample), and fewer who followed a lower educational track (29% vs 43% of total sample).

Overall Procedure of DDP collection

As the sample in this study concerned adolescents, the procedure included steps relevant to studies involving minors. The following five steps will be discussed: (1) privacy

challenges, (2) obtaining parental consent and participant assent, (3) obtaining the Instagram DDPs, (4) processing the DDPs, and (5) coding and counting the DDPs.

Privacy Challenges

As using DDPs for research purposes is an unexplored territory, no clear guidelines for the procedure yet exist. DDPs contain private information as well as information from other platform users who cannot provide consent for participation but are connected to the accounts of participants. Thus, careful consideration regarding privacy and ethical challenges in the process is warranted. In the current study, we carefully examined the data donation process together with a data manager and privacy lawyer of the university prior to the start of the overall study. We paid special attention to the transparency and comprehensibility of the information that was shared with the prospective participants and their parents about the data donation procedure and the acquired content of the DDPs. The DDPs were stored on protected storage drives from the university and the images and text were deidentified. Moreover, the number of researchers who have access to the raw data was minimized and analyses were done on the deidentified data.

Obtaining Parental Consent and Informed Participant Assent

Participants were recruited via a large secondary school in the south of the Netherlands. By means of school visits and follow-up emails before the start of the project, parents of all 745 potential participants (all 8th and 9th graders) received information from the researchers about the full study, including the data donation component. Of the 400 participants who received parental consent, 388 provided assent for the larger study in November 2019. Of these 388 participants, 287 (74%) obtained parental consent for the data download portion of the study. Of these 287 participants, 209 indicated to have an Instagram account and still took part in the larger study in June 2020. These participants received information about (a) what they would be sharing with the researchers if they agreed to

participate, (b) the pseudonymization process with examples of anonymized images, (c) how the data would be stored, and (d) the type of information the researchers were interested in. The reward for their data donation was 5 euros. Of the 209 participants, 148 provided informed assent and automatically proceeded with the data donation process.

Obtaining the Instagram DDPs

After providing informed assent, participants automatically proceeded to an online survey environment and first received questions about how many and what type of accounts they wanted to share, for example regular accounts, fan accounts, or hobby accounts. Next, they received detailed, visual instructions on how to download and share their Instagram DDPs via the Instagram website or the smartphone app. The actual DDP procedure consisted of two main steps: requesting the DDP from Instagram and uploading the DDP to the research drive. At both steps, researchers answered participants' questions and sent reminders if the step was not completed. Of the 148 participants who provided assent, 44 were not willing or able to complete the two steps, due to a set of challenges encountered.

At the first step, participants had to log into their Instagram accounts to request the download and to confirm the email address at which they wanted to receive the data. This turned out to be a first challenge. Some participants did not know their Instagram username and password as they were used to log in automatically on their smartphone Instagram app. These adolescents had to reset their password. Others needed to ask permission from their parents to log into Instagram, and yet others did not know the email address that they had used to create the Instagram account. In our attempt to prepare participants to have this information at hand before starting the data download, some participants thought they had to share the login information with the researchers, making them distrustful of the process.

At the second step, within a maximum of 48 hours after participants' request, Instagram sent the data download zip files to the email address participants had provided.

Some participants received one and others up to 12 zip files, depending on the size of their DDP and the number of Instagram accounts. Once the DDPs arrived, participants needed to download the zip files from their email and upload them to a protected server of the university via a private link. Each participant received their own individualized link to a data folder that was pre-labeled with their participant number. This came with a second set of challenges. Some participants did not have enough space on their phone for the zip files to download, or they could not find the zip files on their phone once it was downloaded. Others were not able to upload complete zip files and instead had to open the zip file and upload individual subfolders and files. Large Instagram accounts consisted of multiple zip files, each of which also had to be uploaded separately. Depending on the individual zip file sizes and internet connection, this uploading process could take a few seconds to several minutes. In some cases, the home or school connection was not sufficiently stable or fast for the files to upload, resulting in files that were not fully uploaded and could not be opened.

In total, 104 participants donated 112 accounts. Of the 44 participants who ended up not sharing their account(s), 30 participants officially indicated that they no longer wanted to donate their account, either due to login troubles (n = 1), failed upload attempts (n = 5), or an unknown reason (n = 24). The other 14 participants simply stopped replying to prompts from the research team.

Processing the DDPs

By means of Python scripts, all DDPs were opened to prepare for deidentification and analyses. Of the 112 accounts, two accounts were removed from further analysis due to technical issues, resulting in a final total of 110 accounts of 102 participants. Of the 110 accounts, 13 accounts were incomplete, because they contained parts that could not be opened, had one or more parts that were missing, and/or had empty parts.

Each unzipped DDP contained folders and text files (JSON). Figure 1 presents an example of the structure of a DDP. The folders contain all the media files (images and videos) that the participant ever shared with others on Instagram. The text files contain information on the activities the participant engaged in, such as the captions of posts and stories, and the content of messages. To illustrate this structure, the “media” text file contains information about every story and post of the account. Within this text file, each post and story update is timestamped and contains a path to the image or video file, and the caption that was created with that post or story (see Figure 2 for an example). In the current study, only the text files were used as they contained all information needed to explore the type and frequency of the Instagram activities adolescents engaged in.

In line with the ethical and privacy agreements, the DDPs were deidentified with a Python script that combined the anonymization and pseudonymization steps. The script removed complete files from the DDPs that contained private information and were not needed for analysis, such as “uploaded contacts” and “devices,” which, respectively, provided contact information of others and listed the devices from which the Instagram account was accessed. In the remaining files, the scripts replaced names, phone numbers, email addresses and URLs with the codes “_name,” “_phone number,” and “_URL,” respectively. In addition, participants’ usernames were deidentified by replacing them with a participant number. Finally, on the videos, faces and text on images that contained usernames were covered (for detailed information on these scripts see Boeschoten et al., 2021).

An important discovery was that the composition of the DDPs depends on the time of donation. Instagram regularly adds functionalities, affecting what is in the DDPs. This may result in either entirely new folders and/or text files or in new information that is distributed across several existing folders and files. Moreover, and importantly, these new folders and text files are typically not rolled out simultaneously across all users. This means that the

structure and content of the DDP is not stable across time and participants. In the current study, the text files and related functionalities “guides,” “fundraisers,” and “events” (see Figure 1) were fairly new and hardly used by participants. We therefore did not include them in our analyses.

Coding and Counting the DDPs

Python scripts were used to explore the (1) overall characteristics of the account (e.g., number of followers, private vs public account, date of creation of account) and (2) activities of the account (e.g., number of posts, stories, likes, and comments) during the eight-month period of the project. Table 1 shows a description of the text files used in this study and the characteristics and activities we obtained from them. Even though providing descriptives of these text files seems a straightforward task, three main challenges arose during the analysis of the DDPs.

The first challenge pertains to decisions of the researchers as to how to assign frequencies to the Instagram activities. For example, one post can contain multiple images or videos. In the DDPs, these components are stored separately, but they have the same timestamp. We decided that all information with the exact same timestamp should be counted as one post. For stories, this is more complex. For example, a participant can post multiple images to one story, but this can only be done one by one and therefore each update is stored with a slightly different timestamp. Because participants sometimes updated their stories several times throughout a day and because it proved impossible to determine what the user had intended as “one story update,” we counted each listing as a separate story post.

This same challenge applied to counting direct messages. We were able to identify and use seven different types of direct messages in the DDPs in our sample, including text messages, GIFs, and stories. These direct messages were each listed as separate messages with their own timestamps. Some messages contained multiple message types, for example an

image with text. We decided to take the same approach as we did with stories and posts and counted each message with its own timestamp as one direct message (see Table 1).

Identifying and operationalizing a chat was even more challenging. Chats are synchronous exchanges of direct messages between two or more fellow users about one or more topics within a certain time frame. However, it proved to be impossible to determine when chats started or ended. We found that some direct messages were exchanged hours apart, but were still about the same topic, which begs the question how to operationalize chatting. Therefore, we counted the number of different chats based on composition of the chat groups and the direct messages that were sent in each group.

A second challenge was the lack of clarity about the Instagram functionality that each text file in the DDP referred to, as the name did not always reveal its exact function. For example, for the text file “seen_content” it is uncertain what counts as “seen”. Based on our assessments, it most likely is the content the participant actively clicked on. As with all text files, it is uncertain to what extent this information is complete or whether more categories of “seen_content” might exist (see Table 1 for our selection). Anyhow, the text file “seen_content” does not provide information on stories seen.

And finally, a third challenge was that the names of the sections do not optimally describe their content. For example, the text file “seen_content” starts with “chaining_seen”, which is a timestamped list of profile names of others, seemingly all profiles a participant has clicked on while browsing. However, at the time of writing it is unclear if these are all or only a selection of profiles the participant ever clicked on.

Results: Preliminary Findings From the Instagram DDPs (Aim 2)

Our second aim was to provide insight in what DDPs can teach us about adolescent Instagram use. We first provide an overview of account characteristics, followed by the activities the adolescents engaged in. The units of analysis are the 110 accounts rather than

the 102 participants. The findings cover the 32-week period running from 21 November 2019 to 1 July 2020. Two accounts had only one of the text files and missed all others, whereas four other accounts missed one text file. In all these cases, we only report the characteristics and activities for the accounts that did have the data files available.

Account Characteristics

Account characteristics are the general characteristics of the account at the moment of the data download, and they include the date at which the Instagram account was created, whether the account was private or public, and the number of connections.

Date Account Creation

The 110 Instagram accounts of this sample were created between September 2013 and June 2020. Six accounts were created during the eight-month period of the study. We kept these accounts as part of the sample as they are illustrative of natural fluctuations in use.

Private or Public Account

At the time of the data donations, 73% of accounts were set to private, whereas the remaining 27% of account were set to public.

Number of Connections

On average, the accounts had 416 followers (range = 0 – 2,098) and followed an average number of 787 accounts (range = 1 – 3,628). In addition, accounts blocked on average 17 accounts of others (range = 0 – 1,170). Finally, 59 accounts had marked 675 connections as close friend, with an average of 6 (range = 0 – 51) close friends across all accounts.

Activities Across the Study Period

The statistics reported in this section represent all activities in the 110 accounts during the 32-week period (Nov 21, 2019 – July 1, 2020), unless they contained missing files.

Connections

As Table 2 shows, on average, accounts gained 91 new followers during the 32-week study period, which implies an average of three new followers a week. But this varied considerably across accounts (range = 0 – 1,929). During this period, accounts also started following on average three accounts per week, which also varied considerably across the period (range = 0 – 3,232). In addition, 86 accounts sent 820 follow requests, which comes down to an average of one per month across all accounts. During this same period, 47 accounts blocked 176 users (range = 0 – 73) and 25 accounts added 194 accounts as close friends (range = 0 – 51).

Posting, Commenting, Liking

As Figure 3 shows, posting was relatively uncommon on Instagram; only 57% of the accounts engaged in it. On average, each account posted less than one photo a month (range = 0 – 148). Posting videos was even less common: No more than 9% of all accounts posted in total 30 videos across the 32-week period. On average, an account commented four times per week on posts of others. Liking of posts of others turned out to be the most popular Instagram activity, with an overwhelming total of 466,249 likes, or on average of 135 a week per account. Finally, 91 accounts liked 13,200 comments of others, leading to an average of around four likes a week across all accounts.

Updating Stories

Updating stories happened more often and by more accounts than posting, 73% of accounts updated their stories 6,381 times. On average, each account updated their story twice a week. Many accounts also interacted with the stories of others by filling out their polls, questions, sliders, quizzes, or countdowns. Specifically, accounts responded to polls around once a week on average.

Direct Messaging

The accounts sent 123,741 direct messages in 4,417 different chat groups, that is, on average 36 direct messages per account per week. Chat groups had on average two members, who sent on average around six messages a week. Most direct messages were text messages (84%), but also posts (11%) and stories of oneself or others (5%) were relatively popular to send in a direct message (see Table 2 for all types of direct messaging).

Seen Content

The accounts clicked on 96,561 profiles, posts, videos, or ads, with an average of 28 a week. Over half (54%) of the content that was clicked on were images posted by other accounts. Noteworthy is that 11% of this content were ads.

Other Instagram Activities

The DDPs also revealed participants' searching, shopping, and saving of content. The searches included hashtags, places, and users, of which the latter was by far the most popular (98% of all searches), occurring roughly once a week per account. In total, product information was viewed 410 times, implying less than once a month across all accounts. Saving posts was relatively popular and happened 43,414 times, or around 13 times a week per account.

Fluctuations of Activities Across Time and Accounts

The frequency of activities fluctuated over time, with, for example, a peak in messages and posts around public and school holidays (i.e., December and May). But not all accounts contributed equally to these activities. Figure 4 shows that a few accounts were dominant in each activity. This is most apparent for posting. For example, 43% of the already infrequent number of posts came from only two accounts. Likewise, 40% of all story updates came from only two accounts. This skewed balance even held for direct messaging: More than half (54%) of all direct messages came from seven accounts, owned by six adolescents.

Although there was some overlap in the accounts that were active across all activities, different accounts dominated different activities. As Figure 5 shows, some accounts engaged

mostly in sending direct messages, whereas others preferred to update their stories, and yet others engaged in all activities fairly equally over the 8-month period.

A final notable trend is that the activity levels did not just vary over time, or across accounts, but also within account over time. As Figure 6 demonstrates, accounts can be extremely active at one point in time, and far less active at other moments in time.

Preliminary Results Comparative Sentiment Analyses

DDPs' main asset is that it provides insight in the content of what participants share on Instagram and how the content shared may differ across Instagram activities (e.g. publicly shared or privately shared). Therefore, we explored the potential of unsupervised automated textual analyses by analyzing the sentiment of a sample of captions of posts and direct messages of the participants in two steps. First, we compared two manual codings of 188 captions with the lexical approach sentiment analysis tool VADER, frequently applied to social media texts (Hutto & Gilbert, 2014). At a second step, we compared two manual codings of 181 direct messages to those with VADER and a second tool used in social media research, SentiStrength (Thelwall, 2017). Both sentiment analysis tools consider the text, emoticons and punctuation when evaluating the valence.

In line with the approach of van Atteveldt et al. (2021), two researchers manually coded the overall valence of the captions and messages by assigning a 1 to a positive, 0 to a neutral, and -1 to a negative post and compared these scores to the compound scores of VADER and/or SentiStrength. When in doubt, for example, when relevant context was missing or when the message was ambiguous in valence, the two researchers agreed to a neutral code.

Regarding the captions of posts, the intercoder reliability was close to acceptable after one practice round for the two manual codings (Kappa = .69), but unacceptably low when comparing these two manual codings with VADER (Kappa = .15 and Kappa = .19). For the

direct messages, the intercoder reliability of the two manual coders was acceptable (Kappa = .70). Again, the intercoder reliability of the manual coders with VADER, but also SentiStrength was unacceptably low (VADER: Kappa = .29 and Kappa = .24; SentiStrength: Kappa = .27 and .33). Also the reliability between VADER and SentiStrength was equally low (Kappa = .27).

The manual coders agreed that of the 188 captions of posts, the majority (66%) were positive in valance, 17% neutral in valence and only 1% were negative in valence. Disagreement on the valence arose for 13% of the captions, and only pertained to differences between neutral and positive or neutral and negative, but never between positive and negative codings. Of the 181 direct messages, the manual coders agreed that around half (54%) were neutral, 23% were positive, and 8% were negative. The coders disagreed on the valence of 16% of the messages, but again this disagreement never pertained to the codings of positive versus negative valence. The discrepancies between the two sentiment analysis tools and the manual codings also largely pertained to neutral and negative/positive codings.

Discussion: Promises and Pitfalls of Instagram Donations

DDPs come with a host of promises. They provide insights in the uses of a social media platform regardless of whether it was accessed via the phone, tablet, or laptop. They contain all user interactions with the platform from the moment the account is created without any researcher intervention. And they provide timestamped information categorized by media activities, such as posting photos, updating stories, and direct messaging. But DDPs also come with several challenges. Because these challenges may be particularly important for colleagues who plan to collect DDPs, we successively discuss the challenges regarding the procedure of data donations, followed by a discussion of the usability of the content of data donations, and finally the implications of our findings for future research.

Procedure

Acquiring DDPs comes with several procedural steps that cannot be avoided and, thus, participants need to be carefully instructed. First, only participants themselves can request their DDPs from their Instagram account. And second, they need to download the DDP once it has arrived in their email and share it with the researcher. The instructions for these steps depend on the size of the participant's DDP, the number of accounts the participant chooses to download and share, and whether the participant requests the DDP via the Instagram app or via a browser. Moreover, the time between the DDP request and receiving the DDP varies significantly (from several minutes to 48 hours). These differences all complicate tailoring the instructions to each individual participant.

In our study, we instructed participants via stepwise online visual instruction. A disadvantage of this online procedure, which was necessary due to the Covid pandemic, was that an additional 30% of participants dropped out partially because they either did not understand the steps, experienced technical issues, or simply felt the process took too long. Moreover, not all DDPs that were shared were successfully uploaded and some of them missed folders or text files. Much of these challenges could be solved with face-to-face or live video instructions. But, ideally each participant would need to be visited twice, once to instruct them how to download the DDPs and once again to upload their DDP to the secured site of the university. Although researchers may simplify the download and upload process in the future, for example, by creating an app that secures direct upload from the phone to the university, successful DDP donation will likely always be heavily dependent on the participants' compliance and technological savviness.

Understanding social media activities gathered via DDPs can be important in its own right. But it may be even more valuable when linked to predictors, mediators/moderators, and/or outcome variables measured via survey or experience sampling method (ESM) designs. However, whatever the design may be, collecting DDPs should logically take place

at the end of a data collection period to link the two data collection methods, which adds to the laboriousness. For example, in the current study, 388 families provided consent to participate in a larger project on the effect of social media use on adolescents' mental health that ran across eight months. However, at the end of the study period, only one out of four participants actually shared their DDPs. This relatively low compliance rate was due to a variety of reasons. For example, not all participants received consent from their parents, not all adolescents had an Instagram account during the study, and not all adolescents were willing or able to follow the steps to download the DDPs. Given that collecting DDPs inherently leads to low compliance rates, intensive recruitment and compliance procedures are even more pivotal than in any other study design.

In our study, we focused on one platform, Instagram. But given that people use on average about five social media platforms to chat with friends and families and/or to present themselves to the broader audience (Waterloo et al., 2018), a focus on a single platform apparently does not suffice. Ideally, to capture participants' actual social media use, studies should collect DDPs of all social media platforms they use. However, a multiplication of DDPs leads to a multiplication of the complexity and laboriousness of the DDP gathering process. Therefore, gathering and analyzing DDPs can best be accomplished by a team of researchers who can collectively reap the rewards of their investments.

Content

An inherent weakness of survey and ESM studies concern their space limitations. Gathering elaborate information on all social media activities, across multiple platforms, and the content that is seen or shared during these activities is not feasible or desirable (Valkenburg et al., in press). Linking DDPs to survey and ESM studies may circumvent this weakness of survey-based studies. DDPs could support self-reports by providing insight in

frequencies and timing of interactions with a variety of social media platforms, and, with a few exceptions such as Snapchat, give access to the content participants engaged with.

The current study has proven the value of DDPs for optimizing strategies regarding the assessment of social media use when employing multiple methods. For example, consistent with many earlier studies among both adolescents and adults that relied on self-reports (Faelens et al., 2019; Frison & Eggermont, 2020) and digital trace data (Marengo et al., 2021) of social media use, we found that posting occurred with such a low frequency that, due to power problems, capturing its effects would be a lucky shot (Valkenburg et al., in press). Although the low frequency of posting may require a general reassessment of how or why it should be studied, it is now clear that it should not take up much-needed space in an ESM or diary study.

Despite its promises, coding the content of DDPs comes with several additional pitfalls. First, a manual to the content of the DDPs is typically absent. It is quite a task to learn what each folder, file, and code in the DDP stands for. Second, the content of the packages is not stable, as new features are added over time, and these features may not be available for all participants at a specific time. Third, although all information is timestamped, DDPs do not provide insight in time spent on specific activities, such as how much time participants spend on crafting an Instagram post or how much time they spend browsing or looking at a specific post. Such information can only be obtained when DDPs are used in combination with other objective methods, such as the *Screenomics* approach (Reeves et al., 2021).

A fourth pitfall of Instagram DDPs is that they do not provide insight in all the content that is seen. It displays what profiles and posts have been clicked on, but not all posts that passed while scrolling, and no information on the stories the participant has seen. Moreover, DDPs provide limited information on feedback by other accounts, such as the likes or comments received by others. The only exception is for direct messages: DDPs contain the

messages that were sent by others. And, if a researcher is interested in overall social media use, all platforms of interest need to be downloaded separately, while not all platforms are suitable to do so. For example, Snapchat only stores meta-data, such as timestamps, but the content itself is not stored and can thus not be accessed. Here again, a combination with other objective methods, such as mobile sensing (Harari et al., 2020) or *Screenomics* (Ram et al., 2019) could be useful. Of course, it is a balancing act between guarding privacy of non-consenting others in capturing content while also having access to how participants interact with content of others.

Fifth, next to the challenges in coding the activities, coding the content of DDPs is equally labor-intensive and challenging. For example, to get insight in text-based platform interaction, the impressive amount of textual data, from direct messages to posts and comments, would require automated textual content analyses. However, the current lexicon-based sentiment analysis tools are not up to that challenge (van Atteveldt et al., 2021). In our study, we coded a sample of the posts and direct messages manually and with VADER and SentiStrength. Similar to van Atteveldt et al. (2021), we were able to quickly reach satisfactory intercoder reliability for the manual codings, but not for the sentiment analysis tools. The low reliability levels for the dictionary approaches seemed to be due to a variety of reasons, such as participants' use of agreed-upon abbreviations, slang, dialect, ambiguity in meaning such as sarcasm, spelling errors, use of expressions and the fact that most textual exchanges were simply too brief to be meaningful without, for example, the visual context such as the posted images, links to websites or GIFs. Hence, reliable analyses of the DDPs require either manual coding or optimization of current automated textual content analyses, preferably in combination with content analyses of the visuals.

Implications for future research

Our DDPs showed that not all activities, and in particular posting, occur with sufficient frequency to be suitable for all types of analyses, and may only be useful to assess between-person (and not within- or person-specific) differences rather than within-person or person-specific effects. As shown by several other studies, only a few participants are highly active while the majority are ‘likers’ and ‘lurkers’ at best (Park & Macy, 2015; van Driel et al., 2019). This is important, as many studies mainly operationalize posting as active social media use. Moreover, our analyses showed that the majority of posts are made by the same few accounts (see Figure 4), sometimes even owned by the same person. Instagram use varies not only in terms of frequency with which users engage in activities, but also in the type of activities they engage in and when they engage in them (see Figure 5 and 6). These varying patterns of use are important to consider when assessing social media use.

Valid coding of the content of social media interactions is crucial when assessing the antecedents and consequences of social media use. However, we have learned that the coding of the supposedly objective DDPs to measure participants’ social media use still requires subjective choices of researchers. For example, assessing the frequency of direct messaging requires a definition of what a message is. In this study, next to text messages we also counted “hearts” sent in response to direct messages of others and sharing GIFS. Similarly, we learned that chatting is challenging to assess as there is no clear boundary indicating when a chat has started or ended, and some participants were part of a chat, but never sent a message to the other member(s) of the chat group.

A final relevant question is what is “public activity” and what is “private activity” when participants can switch continuously between private and public accounts, but also adapt who can see their content. For example, posts can be made from public accounts with a high number of followers, but only made visible to close friends. Likewise, stories can be shared one-to-one, but also with all followers. And, although most group chats consist of two

members, private messages can also be shared in a large group. And importantly, many adolescents own multiple accounts. One account can be for private activity, and the other for public activity. Thus far, limited studies have taken these aspects of social media use into account.

Ultimately, of course, in the future researchers should implement multiple methods to assess a variety of social media uses across multiple platforms and combine these data with measures of mental health. However, next to the obvious feasibility challenges of these types of research endeavors, we should also consider reliable ways to automatically code social media content - images and text - such as obtained with DDPs, and ultimately have sufficiently large samples to apply machine learning and deep learning to effectively study patterns of social media use and mental health.

Conclusion

DDPs come with several clear challenges. But no one can deny that they offer tremendous opportunities to assess the content of social media interaction in a naturalistic environment. We hope that the honest discussion of the pitfalls encountered in the data collection procedure and coding process of DDPs may stimulate future researchers to follow up on these challenges, for example by simplifying the process of sharing and extracting relevant features from DDPs. This exploratory study also laid bare larger challenges regarding the study of social media use and mental health. Optimization in the assessment techniques of social media use should go hand in hand with the optimization of analyzing techniques.

References

Appel, M., Marker, C., & Gnambs, T. (2020). Are social media ruining our lives? A review of meta-analytic evidence. *Review of General Psychology*, *24*(1), 60-74.

Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: a survey of techniques, tools and platforms. *Ai & Society*, *30*(1), 89-116. https://doi.org/10.1007/s00146-014-0549-4

Bayer, J. B., Ellison, N., Schoenebeck, S., Brady, E., & Falk, E. B. (2018). Facebook in context(s): Measuring emotional responses across time and space. *New Media & Society*, *20*(3), 1047-1067. https://doi.org/10.1177/1461444816681522

Beyens, I., Pouwels, J. L., van Driel, I. I., Keijsers, L., & Valkenburg, P. M. (2020). The effect of social media on well-being differs from adolescent to adolescent. *Scientific Reports*, *10*, 10763. https://doi.org/10.1038/s41598-020-67727-7

Boeschoten, L., Ausloos, J., Moeller, J., Araujo, T., & Oberski, D. L. (2020). [Pre-print] Digital trace data collection through data donation. *arXiv preprint arXiv:2011.09851*.

Boeschoten, L., Voorvaart, R., Kaandorp, C., Goorbergh, R. v. d., & De Vos, M. (2021). Automatic de-identification of Data Download Packages. *arXiv preprint arXiv:2105.02175*.

Burnell, K., George, M. J., Kurup, A. R., Underwood, M. K., & Ackerman, R. A. (2021). Associations between self-reports and device-reports of social networking site use: An application of the truth and bias model. *Communication Methods and Measures*. https://doi.org/10.1080/19312458.2021.1918654

Coyne, S. M., Rogers, A. A., Zurcher, J. D., Stockdale, L., & Booth, M. (2020). Does time spent using social media impact mental health? An eight year longitudinal study.

Computers in Human Behavior, *104*, 106160, Article 106160. https://doi.org/10.1016/j.chb.2019.106160

Dienlin, T., & Johannes, N. (2020). The impact of digital technology use on adolescent well-being. *Dialogues in Clinical Neuroscience*, *22*(2), 135-142. https://doi.org/10.31887/DCNS.2020.22.2/tdienlin

Faelens, L., Hoorelbeke, K., Fried, E., De Raedt, R., & Koster, E. H. W. (2019, 2019/02/05/). Negative influences of Facebook use through the lens of network analysis. *Computers in Human Behavior*, *96*, 13-22. https://doi.org/10.1016/j.chb.2019.02.002

Frison, E., & Eggermont, S. (2020). Toward an integrated and differential approach to the relationships between loneliness, different types of Facebook use, and adolescents' depressed mood. *Communication Research*, *47*(5), 701-728. https://doi.org/10.1177/0093650215617506

Griffioen, N., Rooij, M. v., Lichtwarck-Aschoff, A., & Granic, I. (2020). Toward improved methods in social media research. *Technology, Mind, and Behavior*, *1*. https://doi.org/10.1037/tmb0000005

Griffioen, N., Scholten, H., Lichtwarck-Aschoff, A., van Rooij, M., & Granic, I. (2021). Everyone does it—differently: A window into emerging adults' smartphone use. *Humanities and Social Sciences Communications*, *8*, 177. https://doi.org/10.1057/s41599-021-00863-1

Harari, G. M., Müller, S. R., & Gosling, S. D. (2020). Naturalistic assessment of situations using mobile sensing methods. In W. T. A. (Ed.), *The Oxford handbook of psychological situations* (pp. 299-311). Oxford University Press, USA. https://books.google.nl/books?hl=en&lr=&id=lb71DwAAQBAJ&oi=fnd&pg=PA299

&dq=mobile+sensing+review&ots=jqCsH6FsHr&sig=ZitKIdvShVAH4Tt_qalk93iVHa4#v=onepage&q=mobile%20sensing%20review&f=false

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media,

Johannes, N., Nguyen, T.-v., Weinstein, N., & Przybylski, A. K. (2021). Objective, subjective, and accurate reporting of social media use: No evidence that daily social media use correlates with personality traits, motivational states, or well-being. *Technology, Mind, and Behavior*, *2*(2). https://doi.org/10.1037/tmb0000035

Lind, M. N., Byrne, M. L., Wicks, G., Smidt, A. M., & Allen, N. B. (2018). The Effortless Assessment of Risk States (EARS) tool: An interpersonal approach to mobile sensing. *JMIR Ment Health*, *5*(3), e10334. https://doi.org/10.2196/10334

Marengo, D., Montag, C., Sindermann, C., Elhai, J. D., & Settanni, M. (2021, 2021/05/01/). Examining the links between active Facebook use, received likes, self-esteem and happiness: A study using objective social media data. *Telematics and Informatics*, *58*, 101523. https://doi.org/10.1016/j.tele.2020.101523

Odgers, C. L., & Jensen, M. R. (2020, Jan 17). Annual research review: Adolescent mental health in the digital age: facts, fears, and future directions. *Journal of Child Psychology and Psychiatry*, *61*(3), 336-348. https://doi.org/10.1111/jcpp.13190

Orben, A. (2020). Teenagers, screens and social media: A narrative review of reviews and key studies. *Social Psychiatry and Psychiatric Epidemiology*, *55*, 407-414. https://doi.org/10.1007/s00127-019-01825-4

Park, P., & Macy, M. (2015). The paradox of active users. *Big Data & Society, 2*(2), 2053951715606164. https://doi.org/10.1177/2053951715606164

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour, 5*, 1535–1547. https://doi.org/10.1038/s41562-021-01117-5

Ram, N., Yang, X., Cho, M.-J., Brinberg, M., Muirhead, F., Reeves, B., & Robinson, T. N. (2019). Screenomics: A new approach for observing and studying individuals' digital lives. *Journal of Adolescent Research, 35*(1), 16-50. https://doi.org/10.1177/0743558419883362

Reeves, B., Ram, N., Robinson, T. N., Cummings, J. J., Giles, C. L., Pan, J., Chiatti, A., Cho, M., Roehrick, K., Yang, X., Gagneja, A., Brinberg, M., Muise, D., Lu, Y., Luo, M., Fitzgerald, A., & Yeykelis, L. (2021). Screenomics: A framework to capture and analyze personal life experiences and the ways that technology shapes them. *Human–Computer Interaction, 36*(2), 150-201. https://doi.org/10.1080/07370024.2019.1578652

Rideout, V., & Fox, S. (2018). *Digital health practices, social media use, and mental well-being among teens and young adults in the US*. https://www.commonsensemedia.org/

Thelwall, M. (2017). The Heart and soul of the web? Sentiment strength detection in the social web with SentiStrength. In *Cyberemotions* (pp. 119-134). Springer.

Valkenburg, P. M., Meier, A., & Beyens, I. (2022, Aug 18). Social media use and its impact on adolescent mental health: An umbrella review of the evidence. *Current Opinion in Psychology, 44*, 58-68. https://doi.org/10.1016/j.copsyc.2021.08.017

Valkenburg, P. M., van Driel, I. I., & Beyens, I. (in press). The associations of active and passive social media use with well-being: A critical scoping review. *New Media & Society*. https://doi.org/10.31234/osf.io/j6xqz

van Atteveldt, W., van der Velden, M. A. C. G., & Boukes, M. (2021). The validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. *Communication Methods and Measures, 15*(2), 121-140. https://doi.org/10.1080/19312458.2020.1869198

van Driel, I. I., Pouwels, J. L., Beyens, I., Keijsers, L., & Valkenburg, P. M. (2019). *Posting, scrolling, chatting, and Snapping: Youth (14-15) and social media in 2019*. Center for Research on Children, Adolescents, and the Media (CcaM).

Verbeij, T., Pouwels, J. L., Beyens, I., & Valkenburg, P. M. (2021). The accuracy and validity of self-reported social media use measures among adolescents. *Computers in Human Behavior Reports, 3*, 100090. https://doi.org/10.1016/j.chbr.2021.100090

Waterloo, S. F., Baumgartner, S. E., Peter, J., & Valkenburg, P. M. (2018). Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New Media & Society, 20*(5), 1813-1831. https://doi.org/10.1177/1461444817707349

Table 1

Overview of Text Files and Information Obtained From Each File

Name text file	Description of information obtained from each file
Profile	The date at which the account was created. Whether the account is set to private or public.
Connections	We used this file to assess the number of followers, number of accounts followed, number of follow requests sent, number of accounts that are blocked, and the number of accounts that are listed as close friends. We assessed the overall status of these connections at the time of donation (account characteristics) and changes in these connections during the 32-week study period (account activity).
Comments	The number of comments on others' posts.
Likes	Likes of others' posts. Likes of others' comments under posts.
Media	This file contains five sections: "direct messages," "photos," "profile," "videos," and "stories." We assessed the number of photos in direct messages, the number of photos that are posted, the number of videos that are posted, and the number of updates to stories.
Messages	Direct messages are organized in "conversations" or group chats, and each conversation contains all messages that have ever been exchanged with that same group of people. We identified the number of group chats the participant was part of, the number of direct messages that the participant sent in each group chat and across all group chats, different types of messages and focus on: text, images, sharing stories, sharing posts, likes of messages, sharing GIFs, invitation to live video.
Saved	This file consists of two sections, "saved posts" and "saved collections." Saved posts are all posts of others that a participant saved to their account. Saved collections are all collections a participant created under which the posts are saved. This can be compared to Pinterest. The content saved is only visible to the participant and thus not public.
Searches	The number of searches for hashtags, places, and other users.
Seen_Content	The frequency of clicking on content of others. This content is divided in clicks on profiles, photo posts, video posts, and ads seen.
Shopping	Product information viewed and product information saved.
Stories_activities	All interactive story elements of other accounts that a user has participated in. We identified polls, emoji sliders, questions, quizzes, and countdowns.

Table 2

Overview of the Types and Frequencies of Instagram Activities

	Activity	Total frequency	# Accounts involved[1]	*M* (SD)[2]	Range	Median[2]
Connections	New followers	9,853	106/108	91 (194)	0 - 1,929	56
	New accounts followed	19,171	106/108	178 (326)	0 - 3,232	119
	New follow requests sent	820	86/108	8 (13)	0 - 77	3
	New blocked users	176	47/108	2 (7)	0 - 73	0
	New close friends	194	25/108	2 (6)	0 - 51	0
Posting	All posts	624	60/106	3 (14)	0 - 148	0
	Posts with photo	594	57/106	6 (19)	0 - 148	1
	Posts with video	30	10/106	0 (1)	0 - 10	0
Liking	Likes on others' posts	466,249	106/108	4,317 (6,245)	0 - 38,967	1,750
	Likes on others' comments	13,200	91/108	122 (462)	0 - 4,655	25
Commenting	Commenting on others' posts	12,800	102/108	119 (195)	0 - 1,640	63
Updating stories	Story updates	6,381	77/106	60 (202)	0 - 1,337	5
Interacting with stories of others	Polls	4,023	94/108	37 (111)	0 - 1,075	10
	Emoji sliders	636	68/108	6 (14)	0 - 121	1
	Questions	136	38/108	1 (3)	0 - 18	0
	Quizzes	1,151	68/108	11 (20)	0 - 147	3
	Countdowns	179	29/108	2 (10)	0 - 97	0
Direct messaging	All messages sent	123,741	102/108	1,125 (2,585)	0 - 18,316	276
	Chats	4,147	105/108	38 (51)	-	23
	Users in each chat	-	-	2 (2)	-	2
	Messages sent in each chat	-	-	203 (968)	-	10
	Text	103,711	101/108	960 (2387)	0 - 17,545	201
	Sharing a story	5,701	92/108	53 (106)	0 - 758	18
	Sharing a post	13,481	91/108	125 (369)	0 - 3,413	17
	Sharing a photo	2,765	75/106	25 (81)	0 - 666	2
	Liking a message	442	34/108	4 (19)	0 - 169	0
	Sending a GIF	404	30/108	4 (17)	0 - 154	0
	Invite for live video	1	2/108	0 (0)	0 - 2	0
Content seen	All content seen (clicked)[3]	96,561	107/108	878 (850)	0 - 5,362	670
	Profiles seen	13,037	100/108	121 (162)	0 - 824	39
	Posts seen	52,405	105/108	485 (477)	0 - 2,916	361
	Videos seen	20,023	99/108	185 (192)	0 - 1,005	129
	Ads seen	11,096	95/108	103 (104)	0 - 617	71
Searching	Profile searches	2,782	97/108	26 (27)	0 - 116	17
	Hashtag searches	61	25/108	1 (1)	0 - 7	0
	Place searches	5	4/108	5 (0)	0 - 2	0
Shopping	Saved product information	89	12/108	1 (5)	0 - 50	0
	Product information viewed	410	41/108	4 (11)	0 - 73	0
Saving	Saving posts	43,414	97/108	402 (2322)	0 - 22,772	70
	Saving collections	188	33/108	0 (7)	0 - 34	2

[1]The number of accounts that contained information about this activity. Although the total of accounts was 110, two of the accounts missed all information except for # stories, # posts and # images shared in direct messages. For these specific activities four other accounts missed files.
[2]*M* (SD) and median of all accounts of which the data were available
[3]The content that was clicked on and then viewed. The actual number of profiles, posts, videos, ads seen through scrolling is likely much higher.

Figure 1

Folders[*] and Text Files Included in an Instagram Data Donation Package

direct	Bestandsmap
photos	Bestandsmap
profile	Bestandsmap
stories	Bestandsmap
videos	Bestandsmap
autofill	JSON-bestand
comments	JSON-bestand
connections	JSON-bestand
devices	JSON-bestand
events	JSON-bestand
fundraisers	JSON-bestand
guides	JSON-bestand
information_about_you	JSON-bestand
likes	JSON-bestand
media	JSON-bestand
messages	JSON-bestand
saved	JSON-bestand
searches	JSON-bestand
seen_content	JSON-bestand
settings	JSON-bestand
shopping	JSON-bestand
stories_activities	JSON-bestand
uploaded_contacts	JSON-bestand

[*] The first five folders contain all media (video, images) that the participant has shared from the moment the account was created. The remaining files are JSON text files that contain the captions of stories and posts with links to the relevant media files, messages, and other Instagram activities. All information is timestamped.

Figure 2

Part of a Media Text File That Demonstrates its Structure and Content

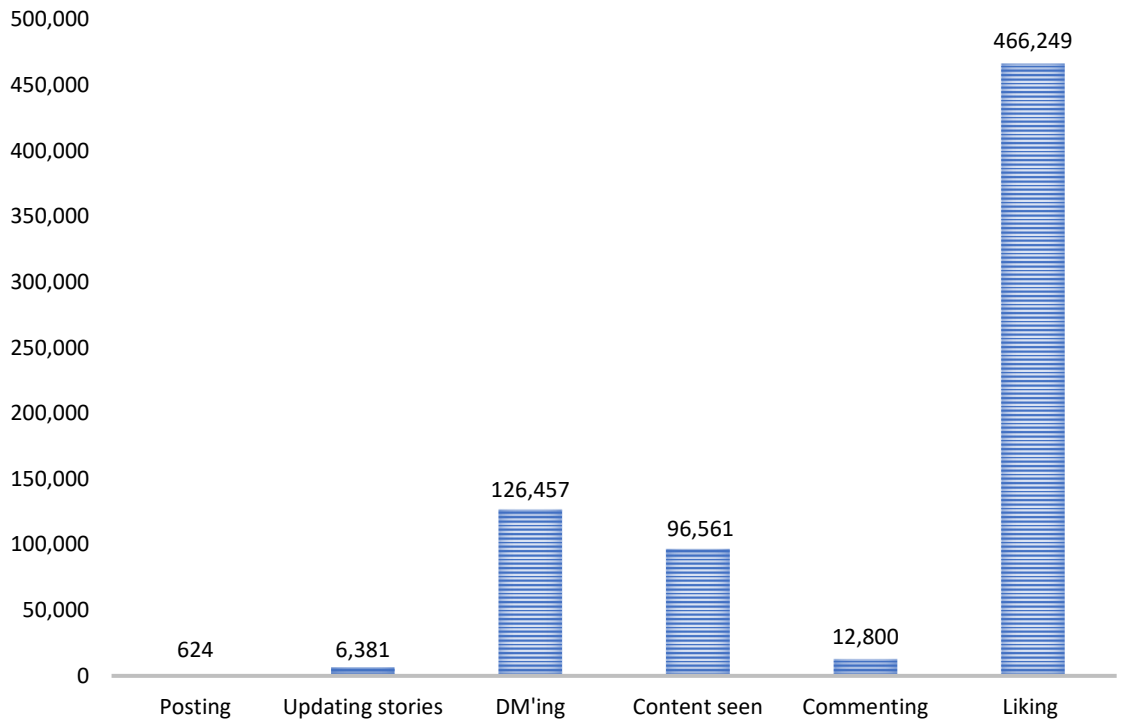

```
{"stories": [{"caption": "Xoxoxo", "taken_at": "2020-10-19T18:26:55+00:00", "path": "stories/202010/863e65e7a09eefd46e816c4830846c54.jpg"}, {"caption": "", "taken_at":
"2020-10-19T10:56:06+00:00", "path": "stories/202010/85fae8fa35067862ed75972efb4da366.mp4"}, {"caption": "", "taken_at": "2020-10-19T10:56:05+00:00", "path":
"stories/202010/c2142737c4c26e7373cf4df8ad98fcb9.mp4"}, {"caption": "", "taken_at": "2020-10-18T19:14:35+00:00", "path":
"stories/202010/8af92f95d0dc90b967ae3c33e1b7206a.jpg"}, {"caption": "", "taken_at": "2020-10-18T19:13:39+00:00", "path":
"stories/202010/206209f0bb867994821b8b402a2f2fcc.mp4"}, {"caption": "Ehm, ok Kim....", "taken_at": "2020-10-18T15:47:30+00:00", "path":
"stories/202010/207dee57ac926b47dd894c5e70565202.mp4"}, {"caption": "", "taken_at": "2020-10-18T15:37:27+00:00", "path":
"stories/202010/6811e529b443ddbf9a71f3ba6066803a.jpg"}, {"caption": "Penelope heeft ook zin in het weekend", "taken_at": "2020-10-17T15:53:39+00:00", "path":
"stories/202010/425203bc01fdba2e1a35cca7e4546270.jpg"}, {"caption": "", "taken_at": "2020-10-17T15:53:04+00:00", "path":
"stories/202010/a84c3bb4ffe88e97f509e3fdf8f2e2ff.jpg"}, {"caption": "Hier is het hele bericht van Beau", "taken_at": "2020-10-17T15:52:46+00:00", "path":
"stories/202010/264840acd8001cf0188b907a9b667019.jpg"}, {"caption": "", "taken_at": "2020-10-16T07:26:33+00:00", "path":
"stories/202010/064142cbee33c41ffd9b1ec323d41e66.jpg"}, {"caption": "", "taken_at": "2020-10-16T07:26:15+00:00", "path":
"stories/202010/1bc16935a95aa5391e423fd715bbccdc.jpg"}, {"caption": "", "taken_at": "2020-10-16T07:25:59+00:00", "path":
"stories/202010/b7b6d0dc2bc9a65e2a8bebb63ec35ea8.jpg"}, {"caption": "", "taken_at": "2020-10-16T07:25:43+00:00", "path":
"stories/202010/60cd33b2a1a3092689a3e3fc45036970.jpg"}, {"caption": "", "taken_at": "2020-10-16T07:25:26+00:00", "path":
"stories/202010/59c4c1658b636f50ae0c3cf8dd6027dd.jpg"}, {"caption": "", "taken_at": "2020-10-15T12:46:46+00:00", "path":
"stories/202010/ae01289a21e0b636669760cab71ded1c.jpg"}, {"caption": "", "taken_at": "2020-10-15T12:43:45+00:00", "path":
"stories/202010/c70ab9a1f275f20eed31f355610b2639.jpg"}, {"caption": "", "taken_at": "2020-10-15T06:39:02+00:00", "path":
"stories/202010/eb1beb1d041e532a10c8a91771c9c821.jpg"}, {"caption": "", "taken_at": "2020-10-15T06:38:16+00:00", "path":
"stories/202010/9629fb52e557d6123a86449b3fbc6847.mp4"}, {"caption": "", "taken_at": "2020-10-14T15:48:35+00:00", "path":
"stories/202010/28e8d45416c3d8115946ea3a70e9105f.jpg"}, {"caption": "", "taken_at": "2020-10-14T15:48:14+00:00", "path":
"stories/202010/9a6985181dd363aa4e118df1b9514bc1.mp4"}, {"caption": "", "taken_at": "2020-10-14T15:47:47+00:00", "path":
"stories/202010/ec2af87e529f475bd48dda83e5de62c5.mp4"}, {"caption": "", "taken_at": "2020-10-14T09:33:09+00:00", "path":
"stories/202010/2b556e267279ff9657a2e70f6ef148a0.mp4"}, {"caption": "", "taken_at": "2020-10-13T17:43:50+00:00", "path":
"stories/202010/234113f7e67fff874610c5a96a3b4396.jpg"}, {"caption": "", "taken_at": "2020-10-13T09:46:29+00:00", "path":
"stories/202010/d6918fe4f4d74def7c4e74e10f28e0a7.jpg"}, {"caption": "", "taken_at": "2020-10-12T20:35:12+00:00", "path":
"stories/202010/a2c159a20c4767076c65edfb1dacbf18.jpg"}, {"caption": "", "taken_at": "2020-10-12T20:34:08+00:00", "path":
"stories/202010/5dbfffea6775a7cc02b432f7ca902bfc.jpg"}, {"caption": "", "taken_at": "2020-10-12T20:33:35+00:00", "path":
"stories/202010/ebccdbaa5a67aee748799497052082ae.jpg"}, {"caption": "", "taken_at": "2020-10-12T20:29:58+00:00", "path":
"stories/202010/c112f0442e99e2df1314b1abe71284c7.mp4"}], "photos": [{"caption": "There are so many cool websites you should check out about bears.\nCheck this photo
collection of bears by photographer Jill Greenberg.\n\nI know Jill very well. You can send her an email when mentioning my name (Bea de Beer) at
Jillthephotographer@outlook.com\nDon't tell her you have it from me, but this is her phone number +31678910111\n\nPlease reply with your own favorite website, email and
phone number so we know you are not just some rando. And in case you are still in this thing.\n\nhttps://www.boredpanda.com/photography-bears-photoshoot-studio-jill-
greenberg/?utm_source=google&utm_medium=organic&utm_campaign=organic\n\n@egelliefhebber\n@snowecho212\n@horsesarecool52 \n@insta4dummy \n@geese_person \n@iliketodance19
\n@kippie_toktok \n\nWie tagged de rest??", "taken_at": "2020-10-20T12:01:04+00:00", "path": "photos/202010/4eb22034c70791a851b32856fdfee2fe.jpg"}, {"caption":
"Allemaal gezichten. #differentfaces @onemillionfaces", "taken_at": "2020-10-19T10:55:24+00:00", "path": "photos/202010/cc9dd3580c44e98ab21aef6c3222728e.jpg"},
{"caption": "Allemaal gezichten. #differentfaces @onemillionfaces", "taken_at": "2020-10-19T10:55:24+00:00", "path":
"photos/202010/2c855e509a802f5f85b483e5b55807c1.jpg"}, {"caption": "Allemaal gezichten. #differentfaces @onemillionfaces", "taken_at": "2020-10-19T10:55:24+00:00",
"path": "photos/202010/617219aa9c542b36eb29256e151af6da.jpg"}, {"caption": "Allemaal gezichten. #differentfaces @onemillionfaces", "taken_at": "2020-10-
19T10:55:24+00:00", "path": "photos/202010/b958658eaea1d1f065834ede098e8892.jpg"}, {"caption": "Allemaal gezichten. #differentfaces @onemillionfaces", "taken_at":
"2020-10-19T10:55:24+00:00", "path": "photos/202010/352bda45f77ee5ed1f091dbb397b8628.jpg"}, {"caption": "Allemaal gezichten. #differentfaces @onemillionfaces",
"taken_at": "2020-10-19T10:55:24+00:00", "path": "photos/202010/9f62816087ef71bce7ebd9cb3a082ae8.jpg"}, {"caption": "Allemaal gezichten. #differentfaces
@onemillionfaces", "taken_at": "2020-10-19T10:55:24+00:00", "path": "photos/202010/fd0c50bd0ffa9db8ae74b55e239501f6.jpg"}, {"caption": "Allemaal gezichten.
#differentfaces @onemillionfaces", "taken_at": "2020-10-19T10:55:24+00:00", "path": "photos/202010/ea6bc904fe393c2054e2a6b7da7902f3.jpg"}, {"caption": "Allemaal
gezichten. #differentfaces @onemillionfaces", "taken_at": "2020-10-19T10:55:24+00:00", "path": "photos/202010/c230a40638bfa92f7d72b3e665a81166.jpg"}, {"caption": "Look
alikes?\nhttps://www.insider.com/celebrities-who-look-alike-2017-1?amp&__twitter_impression=true", "taken_at": "2020-10-18T19:12:27+00:00", "path":
"photos/202010/dbb5daab1890e402b6cef1e4957c31c5.jpg"}, {"caption": "Look alikes?\nhttps://www.insider.com/celebrities-who-look-alike-2017-1?
```


Note. The content displayed starts with the section “stories” and is followed by “photos,” which are the highlighted areas in the image.

This is an account created for research purposes by the researchers of the current study.

Figure 3

*Comparison of Frequencies Across Instagram Activities**

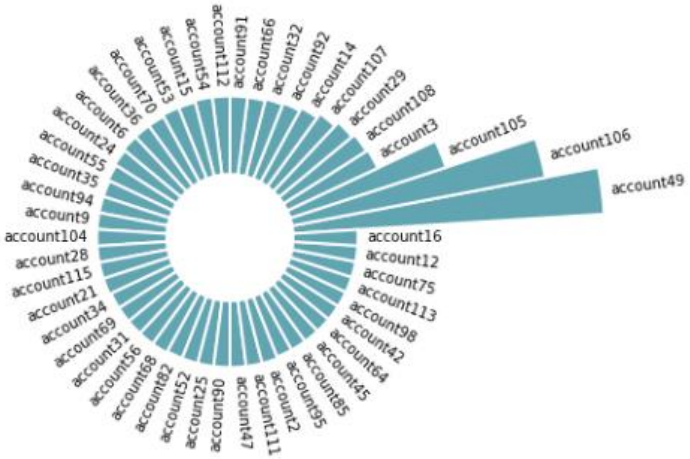


*Posting = posting of photos and videos; Updating stories = all updates to stories; DM'ing = all direct messages sent; Content seen = clicking on profiles, photos, videos, and ads; Commenting = comments on posts; Liking = liking of posts.

Figure 4

Contribution of the 50 most active accounts to the frequency of (a) posting, (b) updating stories, and (c) direct messaging across the 32-week study period

(a) Posting



Note. Range of posting across all accounts = 0 - 148.

(b) Updating stories

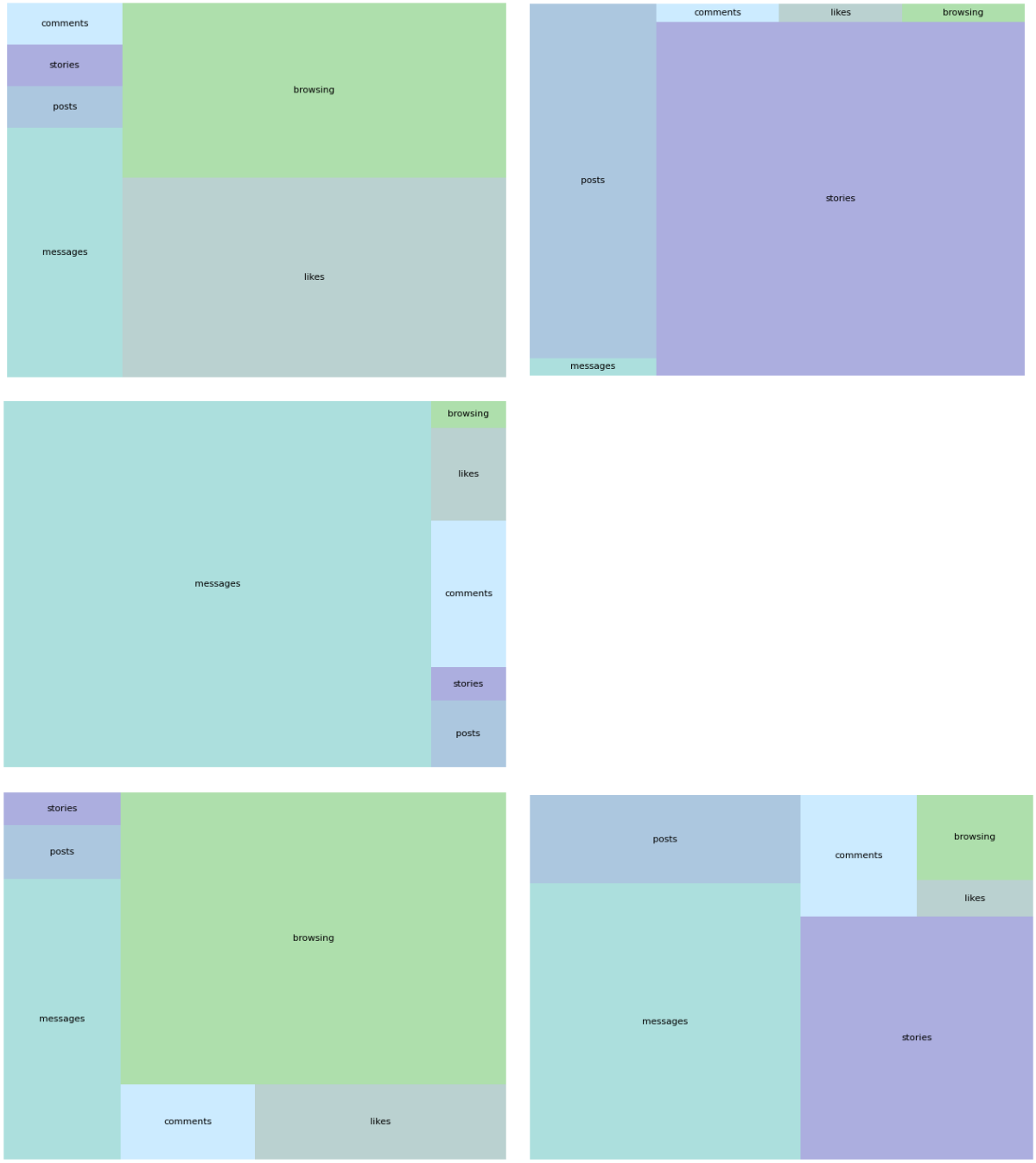
Note. Range of updating stories across all accounts = 0 - 1,337.

(c) Direct messaging

Note. Range of direct messages sent across all accounts = 0 - 18,982.

Figure 5.

Activity Patterns Across Five Randomly Selected Accounts



Note. Posts = posting of photos and videos; Stories = all updates to stories; Messages = all direct messages sent; Browsing = clicking on profiles, photos, videos, and ads; Comments = comments on posts; Likes = liking of posts.

Figure 6

Fluctuations in posting, updating stories, direct messaging, commenting, and liking over time from the top 10 accounts engaging in each activity