ARTICLE

Check for updates

Pandemic in the digital age: analyzing WhatsApp communication behavior before, during, and after the COVID-19 lockdown

Anika Seufert [1✉], Fabian Poignée [1], Tobias Hoßfeld[1] & Michael Seufert [1]

The strict restrictions introduced by the COVID-19 lockdowns, which started from March 2020, changed people's daily lives and habits on many different levels. In this work, we investigate the impact of the lockdown on the communication behavior in the mobile instant messaging application WhatsApp. Our evaluations are based on a large dataset of 2577 private chat histories with 25,378,093 messages from 51,973 users. The analysis of the one-to-one and group conversations confirms that the lockdown severely altered the communication in WhatsApp chats compared to pre-pandemic time ranges. In particular, we observe short-term effects, which caused an increased message frequency in the first lockdown months and a shifted communication activity during the day in March and April 2020. Moreover, we also see long-term effects of the ongoing pandemic situation until February 2021, which indicate a change of communication behavior towards more regular messaging, as well as a persisting change in activity during the day. The results of our work show that even anonymized chat histories can tell us a lot about people's behavior and especially behavioral changes during the COVID-19 pandemic and thus are of great relevance for behavioral researchers. Furthermore, looking at the pandemic from an Internet provider perspective, these insights can be used during the next pandemic, or if the current COVID-19 situation worsens, to adapt communication networks to the changed usage behavior early on and thus avoid network congestion.

[1] University of Würzburg, Würzburg, Germany. ✉email: anika.seufert@uni-wuerzburg.de

Introduction

The spread of SARS-CoV-2, the virus that causes COVID-19, severely affected the daily lives of the human population all around the world. The virus first appeared at the end of 2019 in China and spread around the globe in a few months due to its high infectiousness. The associated disease COVID-19 could turn out as a severe respiratory sickness with high mortality, such that COVID-19 was declared a global pandemic in March 2020 (World Health Organization, 2020). As a reaction, many governments enforced extensive measures to protect their citizens and to control the spread of the virus. Probably one of the most drastic but also most widespread measures was the so-called lockdown, also known as stay-at-home orders or curfews, which was enforced in different countries and territories in varying severity. Some lockdowns just limited the time when people could leave their homes, while others relied on total movement control. By the end of March 2020, over 100 countries worldwide had instituted either a full or partial lockdown, which affected the daily life of billions of people (Dunford et al., 2020).

As the first global lockdown in the digital age, COVID-19 changed people's habits on many different levels. Physical distancing, i.e., reducing human-to-human interactions to slow the spread of the virus, had become commonplace. Note that, although it is commonly also referred to as social distancing, it is not actually meant to describe a limitation of people's social behavior, thus, we prefer the term physical distancing in this work. With physical distancing, almost everything had to happen at home, in particular education, work, entertainment, and also social interactions. Thus, most of the working time as well as the free time took place on the Internet, which resulted in a traffic increase of 80% on media sites (Statista Digital Market Outlook, 2020). The media in-home consumption increased particularly strongly in the area of news (67%), streaming services (51%), television (45%), and mobile instant messaging (MIM) services (45%) (Statista Digital Market Outlook, 2020). While the interest in news was due to the monitoring of the current pandemic situation, the increase in streaming services and television could be attributed to entertainment. Also, MIM applications played an important role to keep people connected during this challenging time, in which it was not possible to meet in person. However, while the effects of COVID-19 have already been studied in related work for many areas, such as environment, psychology, or Internet traffic, little attention has been paid to MIM communication.

In this paper, we therefore investigate how communication behavior in MIM applications has changed as a result of the lockdown, but also as a result of the pandemic as a whole. For this, we consider the most successful MIM application WhatsApp, which was the number one messaging application during the lockdown, generating the highest amount of MIM network traffic worldwide (Sandvine, 2020b). In particular, we answer the following research questions:

RQ1: Has WhatsApp communication behavior changed during the COVID-19 lockdown in March/April 2020?

RQ2: Are there lasting effects of the pandemic on communication behavior in WhatsApp?

Even before the pandemic, 40% of all messaging traffic on mobile devices was caused by WhatsApp (Sandvine, 2020a). Thus, WhatsApp is well suited to study the effect of the lockdown and the pandemic on MIM communication behavior.

To answer the research questions, we utilize a large data set of 2577 private WhatsApp chats, including 25,378,093 anonymized messages from 51,973 anonymized WhatsApp users, which was collected using the tool *WhatsAnalyzer* (Schwind & Seufert, 2018). Using this data set, we evaluate the influence of COVID-19 and all resulting restrictions on WhatsApp usage behavior. First, to address *RQ1*, we focus on the direct effect of the strict lockdowns by comparing the communication behavior in March and April 2020 with the same period in 2019 in terms of number and type of sent messages, as well as communication activity during the day. Second, to answer *RQ2*, we analyze the long-term influence of the COVID-19 pandemic on WhatsApp communication in an entire pandemic year compared to the pre-pandemic year. Our results confirm that the lockdown drastically changed the communication behavior in WhatsApp. While some of these effects were only visible for a short time, others persisted even after the lockdown, which indicates that the COVID-19 pandemic has a long-lasting influence of people's daily lives. We think that the results of our work are of great relevance for behavioral researchers as they show that even anonymized chat histories can tell us a lot about people's behavior and especially behavioral changes during the COVID-19 pandemic. Furthermore, looking at the pandemic from an Internet provider perspective, these insights can be used during the next pandemic, or if the current COVID-19 situation worsens, to adapt communication networks to the changed usage behavior early on and thus avoid network congestion.

The remainder of this work is structured as follows. Section "Related work" provides related work on the communication behavior in mobile messaging applications as well as on the influence of COVID-19 and the strict lockdown on different areas. In Section "Collecting WhatsApp chats histories", the methodology of collecting private WhatsApp chats using the collection tool is presented. Next, in Section "Effect of the lockdown", we evaluate the influence of the COVID-19 lockdown on user communication behavior while in Section "Lasting Effect of the Pandemic" the effect on the whole pandemic year is investigated. Finally, Section "Conclusion" concludes this work.

Related work

MIM is one of the most used communication methods today. Thus, several works exist on how people communicate with each other and how this communication has been changing in the last decade due to MIM applications, such as WhatsApp, Facebook Messenger, and WeChat, which are currently the most popular MIM applications (Tankovska, 2021). While the effect of the COVID-19 pandemic on many different areas of life was already investigated, the influence on the communication behavior in MIM application has been neglected so far. In the following, related works are briefly summarized by first focusing on research about the usage behavior in MIM applications, and afterwards, by presenting related work investigating the influence of COVID-19 on Internet and MIM use.

Usage behavior in MIM applications. Nowadays, it is impossible to imagine our daily communication without MIM applications and in some regions, such as Europe, it seems that MIM has replaced traditional SMS (Pinchas, 2018). In this respect, Church & de Oliveira (2013) compared the communication behavior in WhatsApp with SMS communication by conducting interviews and a survey. It was found that neither technology was a substitute for the other. However, the cost of SMS had a significant impact on frequency of use and was therefore one of the main reasons for switching to MIM apps, along with social influence. Overall, SMS was seen as more privacy-friendly and formal, while WhatsApp messages tended to be used for more social and informal communications. Further research in the field of

WhatsApp was done by Seufert et al. (2015) and Seufert et al. (2016). Here, the authors focused on the communication within group chats. Therefore, they conducted a survey and collected private chat histories of 243 users to obtain first statistics of group communication and modeled their behavior. The number of messages per WhatsApp group and per user, the content and language of messages, as well as, for the most active groups, the number of messages per day was investigated by Garimella & Tyson (2018) using a data set of public WhatsApp group communication consisting of 178 public groups. In the work of Christian et al. (2015), the communication behavior of more than 2400 WhatsApp users in Germany was evaluated over 1 month. The authors found that WhatsApp was used for more than half an hour per day, which accounted for around 20% of the smartphone usage time.

Apart from the focus on WhatsApp, there is other research in the area of MIM applications and communication within. The offline face-to-face interaction was compared with MIM communication by Knop et al. (2016). Their results show that offline communication is preferred for sharing personal information while the frequency of communication is significantly higher in MIM. The influence of using MIM on the way people coordinate everyday life was evaluated by Ling & Lai (2016). For this, they questioned five focus groups in Singapore and Taiwan with a total of 47 participants. The responses indicated that MIM apps, especially WhatsApp, are commonly used for coordinating social as well as working life.

Internet and MIM use during the COVID-19 pandemic. The spread of COVID-19 is the first pandemic to occur in today's fully globalized and networked society. Even before the outbreak of COVID-19, people were used to using digital media to get information, stay in touch and participate in society. However, the pandemic has significantly changed the way we use the Internet and especially MIM applications, as it could be used as a window to the world during the lockdowns. This can be seen, for example, in the increased Internet traffic. Owing to the fact that a lot of work was done from home during the lockdown and that leisure time also had to take place while maintaining physical distance, COVID-19 strongly influenced Internet traffic. This has been investigated in numerous studies, for example, by Feldmann et al. (2021), Lutu et al. (2020), Boettger et al. (2020), and Xia et al. (2021). Feldmann et al. (2021) used data from a diverse set of vantage points including one Internet service provider, three Internet exchange points, and one metropolitan educational network to evaluate the effect of global lockdowns on traffic shifts. They showed that the traffic volume increased by 15–20% almost within a week in the beginning of the lockdown. Furthermore, a shift of traffic patterns was visible, e.g., during the day or on different days of a week. For example, with the pandemic lockdown in March, the workday traffic pattern shifted towards a continuous weekend-like pattern of the pre-pandemic period. This significant traffic increase was also measured by Xia et al. (2021). Here, Facebook's user-facing traffic from data center regions to point-of-presence sites in four geographical regions rose significantly as they measured a traffic surge of 86% in Asia, 78% in Europe, 65% in North America, and 70% in South America during lockdown. Looking at mobile network traffic, the situation is different, as was shown by Lutu et al. (2020). Using a data set of a large commercial mobile network operator in the UK, they found that the overall download traffic decreased by 20%, especially in densely populated urban areas where it even decreased by 60%. In contrast, the mobile uplink traffic in suburbanites increased by 10% as well as the voice traffic, which increased up to 150%. By evaluating the traffic of Facebook's edge

network serving multiple large social networks, Boettger et al. (2020) showed that changes in traffic demand, user behavior, and user experience over the course of the COVID-19 pandemic were present. While an increase in the popularity of live streaming and messaging applications was accompanied by a significant increase in traffic for these products, the largest impact on traffic resulted from relatively lower traffic growth for video products.

However, it's not just the increased Internet traffic that shows that messaging applications were an important part of many people's lives during the COVID-19 pandemic. As schools were also closed due to the numerous lockdowns, new ways of communication had to be found here as well. Multiple studies from several countries, e.g., by Mawarni et al. (2020), Susilawati et al. (2020), Setiawan & Iasha (2020), and Khan (2021), showed that online learning with the help of the MIM application WhatsApp proved to be an extremely effective method. Other studies by Li et al. (2021) and Amosun et al. (2021) also showed that MIM applications are used to help people navigate times of physical distancing to tackle loneliness and increase well-being.

We have seen that the popularity of messaging applications increased during the COVID-19 pandemic, but there is also an influence of the pandemic on people's communication behavior, i.e., the way people communicate with each other. According to Nabity-Grover et al. (2020), the pandemic has changed the boundary of how and what individuals reveal about themselves on social media. They argued that many people have shared personal information on social media, such as health status and preventive behaviors, e.g., wearing masks and purchasing disinfectants, because they find that sharing such information contributes to the public good. The interaction patterns of individual influencers on the social media networks Instagram and Facebook were investigated by Trevisan et al. (2021). By analyzing and comparing engagement on the posts of these influencers and evaluating the user activity, they showed that there were changes in usage patterns before and during the lockdown, showing an increase in activity and considerable daily and weekly variation.

However, not only the communication behavior in social networks, but also in MIM applications has changed as a result of the COVID-19 pandemic, but there are hardly any studies on this. Ohme et al. (2020) investigated smartphone news use and MIM behaviors of Flemish adults during the begin of the pandemic in February and March 2020. Using a smartphone logging tool for Android phones they found that in March 2020, the average numbers of communication via text messages and messaging apps increased substantially compared to the pre-pandemic month February. However, to the best of our knowledge, no study has investigated yet in detail how the communication behavior in MIM applications has changed during the lockdown, and how these changes have evolved after the lockdown.

The presented related works already gave interesting insights in both, usage behavior in MIM applications and effects of COVID-19 on Internet and MIM usage, showing how important MIM applications were to people during the COVID-19 pandemic. However, an in-depth analysis of the impact of the COVID-19 pandemic and the resulting lockdown on the communication behavior in MIM applications is still missing. To overcome this issue, in this work, we compare the way people communicate using MIM applications before and during the pandemic using a large data set including 2577 private WhatsApp chats.

Collecting WhatsApp chats histories
To obtain insights into the communication behavior in private WhatsApp groups, a large data set of real chat histories is

```
09/30/21, 17:29:12 - Alice created group "Friends"    2021-09-30 17:29:12, User1, <created group>
09/30/21, 17:30:58 - Alice added you                   2021-09-30 17:30:58, You, <added>
09/30/21, 17:32:43 - Bob: Hi Carol                     2021-09-30 17:32:43, User2, 8 chars , 0 emojis
09/30/21, 17:57:08 - +12 1234 12345678: Hey! 😊        2021-09-30 17:57:08, User3, 5 chars , 1 emojis
09/31/21, 12:43:32 - Carol: Hello everybody            2021-09-31 12:43:32, User4, 15 chars , 0 emojis
10/01/21, 08:29:51 - Alice: <video omitted>            2021-10-01 08:29:51, User1, <video>
```

Fig. 1 Screenshot of a chat history as exported by WhatsApp (left), and its corresponding anonymized and normalized version (right). In the anonymization process, user names as well as the content of the message is removed and only relevant information about the communication behavior is kept, like, for example, the message type or the number of characters and emojis.

necessary. To collect these chat histories, the tool *WhatsAnalyzer* (Schwind & Seufert, 2018) was used. This tool is a free to use application for collecting and automatically evaluating WhatsApp chats. It is based on a built-in feature in WhatsApp, which allows to send an email with a text file containing the chat history. With this feature, users can send a chat to our tool, where a unique identifier is generated, system messages are detected, and the remaining chat is anonymized to protect user privacy. The unique identifier is based on a hash of the chat name and the email address of the sender, which allows to easily detect whether a chat was sent multiple times. To reliably detect system messages, we supported only four languages: English, Spanish, German, and Italian. However, users could change the system language of their phone to one of these languages to obtain a compliant export format.

As can be seen in Fig. 1, anonymization means that user names, phone numbers, and message content are removed, and only relevant information about the communication behavior is kept. Moreover, a normalized version of the chat is stored for further analysis, because export formats of WhatsApp greatly vary due to system language, operating system, and/or WhatsApp version. Finally, some basic statistic evaluations of the communication within that chat are calculated, e.g., the total number of messages, the length of the shortest and longest message, and the temporal distribution of all messages. As an incentive to use our tool, these evaluations are sent back to the users, after they sent their chat, to also provide them some analyses of their own communication behavior. Note that users are, of course, informed on our website that the anonymized chats are further analyzed for research purposes.

All collected chat histories were sent voluntarily by people who were interested in the resulting evaluations. This means that no targeted study or recruitment was made. Our tool was only advertised using a web page and social media and can easily be found in popular search engines by people who want to get insights of their WhatsApp chats. This tool in its current form was published in 2017 and is well received as its database already contains 7312 entries and continues to grow. After filtering out incomplete data or chats, which were sent multiple times, 5891 chats remain with 116,441 different users and 75,910,808 messages. Since we investigate specific periods in our evaluation in this paper, trimming the data set for these periods yields a subset of 2577 private chat histories with 25,378,093 messages from 51,973 users. Although no information about the senders exists, the collected data set allows to gain unprecedented, detailed insights into the communication behavior in private WhatsApp chat groups.

Effect of the lockdown
The worldwide lockdown in response to COVID-19 has restricted billions of people, and thus, also changed their behavior on various levels. Therefore, the question arises (*RQ1*) if and to what extent the lockdown has also changed the way we communicate with each other using MIM applications. Thus, in this section, we compare messaging behavior during the lockdown period 2020 with the same period in 2019 using the collected WhatsApp chats.
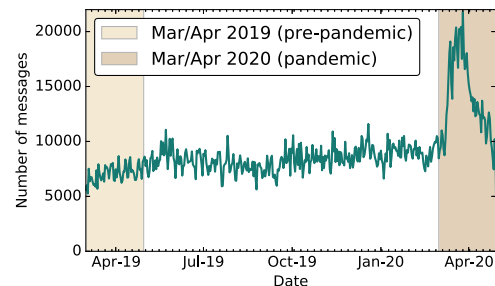

Fig. 2 Number of messages per day. The figure shows the total number of messages for each day from the beginning of March 2019 until the end of April 2020. On the left-hand side the background is colored for the Mar/Apr 2019 pre-pandemic period, on the right-hand side the lockdown period in Mar/Apr 2020 is highlighted.

We define the lockdown start to be the 1st of March 2020 and the end to be the 30th of April 2020. While there was no uniform start of the lockdown worldwide, during these 2 months Mar/Apr 2020 media coverage for COVID-19 spread world-wide and first governmental actions took place (Dunford et al., 2020).

For our evaluations, we consider just the small subset of chats, which were active over the complete span of both periods. Here, active means that the conversation was started before March 2019, and contains messages in both periods. This allows a paired comparison of the pre-pandemic months Mar/Apr 2019 with the lockdown months Mar/Apr 2020 within the same communication group. The total number of considered chats is 345, of which 250 are group chats with more than two active users. The total number of active users within these periods is 8150 for the 2019 2-month period. Note that the number of different senders drops by 46% for the 2020 lockdown period, resulting in 4378 users. Reasons for this could be that users left the group chat or remained inactive during this period, as our data only shows users who actively participated in the group's conversation.

Owing to the design of the tool and the fact that we did not conduct a targeted study with recruited participants, we unfortunately do not have any information about the senders or their countries of origin, and therefore, cannot make any statements about the differences between individual countries. Nevertheless, in March and April 2020, there was a lockdown in virtually every country in the world, so in our opinion, a distinction by country is not necessary here to obtain insightful results.

Message frequency. To check whether a change in communication behavior has occurred as a result of the lockdown, Fig. 2 shows the total number of messages for each day from the beginning of March 2019 until the end of April 2020. On the left-hand side the background is colored for the Mar/Apr 2019 pre-pandemic period, on the right-hand side the lockdown period in Mar/Apr 2020 is highlighted. During the first period, the number of messages never exceeds 9000 messages per day and averages at 7087 messages. Between the periods, the fluctuation stays roughly between 5000 and 12,000 messages per day, having an average of 8429. However, in the lockdown period 2020 there is a massive
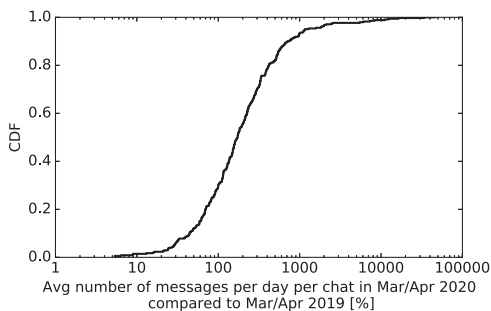
Fig. 3 Difference in average number of messages per day per chat. The figure shows the CDF of the percentage of the average number of messages per day per chat from the Mar/Apr 2020 lockdown period compared to the Mar/Apr 2019 pre-pandemic period.
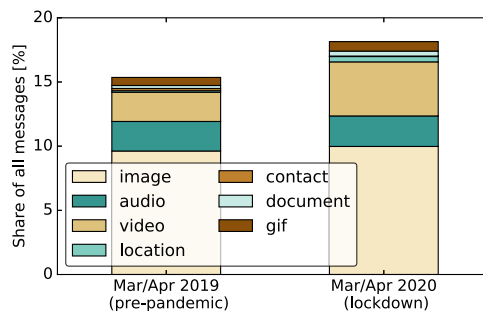

Fig. 4 Share of media for each period. The figure shows the share of sent media messages (e.g., images, video, and gifs) in the pre-pandemic period (left) compared to the pandemic period (right).

spike with a maximum at 21,785 messages per day, which is 2.58 times the average of the Mar/Apr 2019 period. This results in an average number of messages per day of 13,787 during the lockdown period. Towards the end of the lockdown period, the message frequency drops to previous levels of activity. Looking at the total sum of messages per period, the total number of messages almost doubles from 432,348 to 827,241, even though the number of different senders drops from one period to the next. Thus, it can be seen that a significant increase in the total number of messages has taken place within the same WhatsApp chats.

The next question that arises is whether this sharp increase is perhaps only due to certain chats, or whether the statement can be generalized. Figure 3 shows the cumulative distribution function (CDF) of the percentage of the average number of messages per day per chat from the Mar/Apr 2020 lockdown period compared to the Mar/Apr 2019 pre-pandemic period. For each chat, this percentage is calculated as follows:

$$\frac{\overline{n}_{\text{Mar/Apr2020(lockdown)}}}{\overline{n}_{\text{Mar/Apr2019(pre-pandemic)}}},$$

where $\overline{n}_x$ is the average number of messages per day for the given period x. Thus, there is one sample per chat that shows the percentage of communication for the Mar/Apr 2020 lockdown period in comparison to the Mar/Apr 2019 pre-pandemic period. Here, a value of 100% means that the number of messages is the same in both periods. The Kolmogorov–Smirnov test for the distribution of each period gives a p-value < 0.001, so we reject the null hypothesis, which states that both periods have the same underlying distribution. In all, 29% of the chats have a value smaller than 100%, which means that in these chats, less messages were sent in the Mar/Apr 2020 lockdown period than before. The strongest decrease in messages was a reduction to 5.35% of the messages in the 2019 period. This is a reduction by a factor of almost 20. This is expected as conversations in WhatsApp chats can become less active over time. However, for most of the chats the activity increased significantly. In 50% of the chats, the percentage is above 173%, which means that activity increased by more than 73%. For more than 45% of the chats, the activity at least doubled having a percentage above 200%. The maximum is a 454-fold increase for a single data point. It becomes clear that the change in message activity is not caused by single instances but by the majority of chats.

While not only more messages were sent, the proportion of media sent, such as pictures or videos, also changes from one period to the next. The Chi-square test for the frequency of the observed message types in the different periods yields a p-value < 0.001, thus, we reject the hypothesis that both samples have the same underlying distribution. Furthermore, a paired t-test is conducted for each media type where each chat constitutes

a pair that is formed from its share of messages during the lockdown period 2020 and its counterpart in 2019. We apply Bonferroni correction to account for the multiple comparisons problem and reject the null hypotheses that the underlying distributions are the same with a p-value < 0.001. These changes in the share of media are shown in Fig. 4. Here, the share of each media type for both periods is displayed. In total, there are still more text messages sent than media but the share of media messages in all messages increased by more than 18% from the 2019 period (15.35%) up to 18.14% of messages for the pandemic period. There's a slight increase in images as well as audio messages. The share of video increases from 2.27% to 4.21% and is one of the main factors for the increased overall media share. The share of location messages is barely visible for the 2019 period but increases almost fourfold for the 2020 period. A reason for that could be that due to government restrictions users were not able to meet at known locations, but rather met in new, less populated places and used location messages to find each other. The share of contact messages is the only one decreasing from one period to the next, which could be due to the fact that social restrictions prevented meeting new people. Moreover, the share of documents increased from 0.23% to 0.39% of all messages, which could be an indicator for an increased professional or educational use of WhatsApp.

Communication activity throughout the day. An important metric for usage behavior of a communication service is the usage throughout the day. Figure 5 (a) shows the share of messages per hour in the Mar/Apr 2019 pre-pandemic period. It should be noted that all chats are evaluated in their local time. Without government restrictions in daily life, the usage of WhatsApp increases throughout the day and peaks in the late evening. While there is little usage during night-time, in the morning hours the share of messages increases beyond 5% and peaks in the evening (9 pm) at more than 7% of all messages. During the lockdown period in 2020 this behavior changes significantly. Figure 5 (b) shows the difference in time of day usage compared to the Mar/Apr 2019 period. A Chi-square test resulting in a p-value < 0.001 rejects that the observed time of day for both periods belong to the same distribution, and thus, indicates that the time of day usage is significantly different in both periods. There are three major characteristics visible in the figure: The first one is an increase in night-time messaging. During the lockdown period, from 12 am to 3 am increases are visible ranging from 0.27% to 0.77%. A reason for this could be psychological side-effects of the lockdown restrictions, which have been associated with lower sleep quality (Casagrande et al., 2021; Pinto et al., 2020). The second one is an increase of messages around the middle of the day, starting at 8 am and lasting until 3 pm. A maximum increase of 1.45% is visible at 11 am. As a result, the mid-day hours
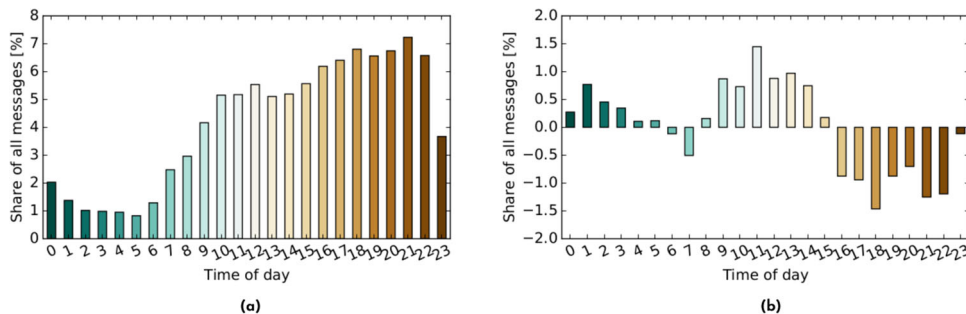
Fig. 5 Share of messages per time of day for the 2-month period 2019 and its changes in the lockdown period 2020. In (**a**), the bars show the percentage of messages per time of day in the 2-month period 2019. In contrast, (**b**) illustrates the changes of the communication behavior from the pre-pandemic period to the lockdown period 2020. Here, each bar shows the differences between the number of sent messages per time of day of both periods.

represent peak usage for the Mar/Apr 2020 lockdown period. A possible explanation for this could be an increased usage of WhatsApp in professional working environments, home offices, and the use of MIM application for school purposes. This overall trend is in line with findings by Feldmann et al. (2021) who reported an increase in mid-day Internet traffic. Third, the increase of messages in the previous two areas comes at the cost of a decrease in the evening hours of 0.70% to 1.46% in the time from 4 pm to 10 pm. A possible explanation could be that as people during the lockdown cannot attend social meetings and other activities, messaging friends or family to arrange appointments may occur less frequently.

Our results show that the communication behavior in WhatsApp significantly changed, which answers RQ1. Focusing on the frequency of messages, we found that significantly more messages are sent as the total number of messages is almost doubled from Mar/Apr 2019 pre-pandemic period to Mar/Apr 2020 lockdown period. In addition, the lockdown period 2020 shows a massive spike, which is more than double of the average of the pre-pandemic period. Furthermore, the proportion of sent media files also changes from one period to the next, as the share of media messages in all messages increased by more than 18%. Considering the activity throughout the day, we showed that the main communication times have shifted away from the evening hours and towards mid-day, leading to a maximum increase of 1.45% at 11 am and a decrease of up to 1.46% at 6 pm in the evening. Furthermore, more messages were sent after midnight from 12 am to 3 am compared to the Mar/Apr pre-pandemic period.

Lasting effect of the pandemic
Even after the hard lockdown period, there were still severe restrictions for many people, so that the entire year from March 2020 to March 2021 was dominated by concerns about COVID-19 and its consequences. So, the question (RQ2) arises whether communication behavior changed not only during the hard lockdown, but changes also could be observed during the full year from the start of the pandemic. Thus, in the following, we analyze changes in the communication behavior in WhatsApp when comparing the entire pandemic year to the year before when there was no pandemic. Furthermore, this allows to determine which effects of the lockdown are temporal and which persist for a longer time.

For our evaluation, we include all chats that are collected from March 2019. The conversations were split into two periods: Messages that were sent in the pandemic year (March 2020 until February 2021) or messages sent in the pre-pandemic year before (March 2019 until February 2020). In contrast to our previous evaluation, here, some chats only appear in one of the two
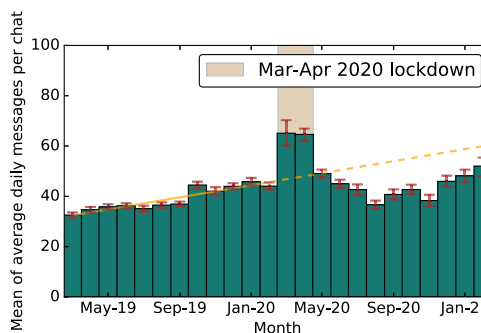

Fig. 6 Average number of the mean of daily messages per chat per month. In this Figure, each bar shows the mean number of daily messages per chat per month with their corresponding 95% confidence intervals on top (red). The trend of the pre-pandemic period (linear regression of the means from March 2019 to February 2020) is marked as dashed yellow line, while the lockdown period (March to April 2020) is highlighted in brown.

periods. One reason for this is that some chats were collected by the tool before the begin of the pandemic and thus, only include messages of the pre-pandemic period. Thus, the data density decreases over time. For the pre-pandemic year from March 2019 to February 2020, our data set includes 2039 different chats. Of those, 64% are group chats. In total, more than 15 million messages by more than 38,000 users are analyzed for this period. For the pandemic year from March 2020 onward, our data density decreases by 29–39% with respect to the total number of tracked chats, unique users, and messages. Nevertheless, the pandemic period still contains 9.7 million messages from about 27,200 users. Since the number of chats in the pre-pandemic and in the pandemic year differ significantly, we limit our evaluation to relative statements as the comparison of absolute values (e.g., the total number of messages) are not meaningful in this case.

As mentioned before, we had no insights into the users' country of origin and thus were not able to compare the impact of individual countries' preventive measures during the pandemic year on communication behavior. Nevertheless, we assume a stable distribution of users' country of origin in our data over time. Thus, we compare both time periods based on our data, which still allows to determine the lasting impact of the pandemic on WhatsApp messaging behavior regardless of locally or temporarily different preventive measures.

Message frequency. To examine whether the pandemic has caused a lasting change in the usage frequency of WhatsApp beyond the lockdown, Fig. 6 shows the mean number of daily

messages per chat, averaged per month with the corresponding 95% confidence intervals. The evaluation ranges from one year before the lockdown (March 2019) to one year after the begin of the lockdown (February 2021). The yellow line shows the linear regression of the mean values per month of the pre-pandemic year, where we can clearly see the trend of an increasing number of messages per day. As already discussed in the previous section, the Mar/Apr 2020 lockdown (highlighted in brown) stands out with a significantly increased number of messages per day. After this period, the average number of daily sent messages drops again and falls below the trend of the previous year (dashed line). Nevertheless, the number of messages is still above the respective months of the pre-pandemic year. Towards the end of the period, the average number of messages per day rises again and eventually approaches the trend of the yellow line. If we compare the average value of the pre-pandemic period with that of the pandemic period (excluding the lockdown peak), the number of messages increased noticeably from 41.02 up to 50.65. Overall, it is clear that the pandemic, and especially the lockdown period, had a strong impact on the number of messages per day, but also that communication behavior stabilized again as expected towards the end of the pandemic year.

Similar to our previous evaluations of the 2-month periods, we also look at the share of media for each 1-year period in Fig. 7. Again, the Chi-square test rejects the hypothesis of the same underlying distribution with a p-value < 0.001. Furthermore, we tested for each media type with a paired t-test and apply Bonferroni correction. Since the number of chats is not the same for both periods, we paired based on the days. We reject the null hypotheses with a p-value < 0.05. In contrast to our previous evaluation in Fig. 4, in which the media usage changed drastically between the Mar/Apr 2019 pre-pandemic period and the Mar/Apr 2020 lockdown period, the overall media usage is roughly the same for both years, but the composition is slightly changed. However, there is an increase in the proportion of users who never send any media at all in the pandemic year. Messages containing images decrease by 0.34%, audio remains almost constant, while video increases by 0.14%. A more significant change can be observed in the exchange of locations, which drops from 1.69% to 0.75% when comparing both periods. The number of contacts is affected even more and decreases from 0.06% to only 0.01% of all messages. A reason for this drop in contact exchanges could be that people have a reduced willingness to meet new contacts, as the possibility of an infection still persists after the initial lockdown period. In contrast, messages with documents increase by 36% up to 0.50% of all messages. This is another sign for increased professional information exchange compared to the pre-pandemic year. Comparing the behavior of the 1-year periods to the behavior seen in the 2-month periods in Fig. 4, there is a normalization in image, audio, and video usage when looking at the full-year pandemic period. In contrast to the
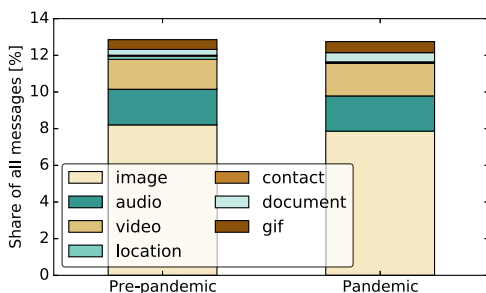
2-month periods, the number of location exchanges decreases. The strongest relative changes for both comparisons are the drastic decrease in contact messages and the rise in documents. Nevertheless, the increase in media messages as well as the drastic change in the distribution of media during the lockdown is shown to have more or less equalized over the period of a full year.

Communication activity throughout the day. In contrast to the media usage, the change in usage over the course of a day persists for the one-year pandemic period. The three patterns of more night-time and mid-day usage with a decrease in the evening that we found in Fig. 5 can be identified as well although they flatten slightly. The night-time pattern persists with an increase of 0.1% to 0.2% for the pandemic year. The peak value for mid-day message growth is again at 11 am with a value of 0.89%. Furthermore, there's a corresponding decrease in evening messaging in the time from 4 pm to midnight for the 1-year periods as well. This is evidence for a lasting change of user behavior and a worrying development when linking the increased night-time activity with sleep disturbances (Casagrande et al., 2021; Pinto et al., 2020).

In general, messaging behavior can vary during the lifetime of a chat due to exceptional events or conversations that can cause a substantial increase in the number of messages per day. Figure 8 shows the share of days for each chat where the number of messages was exceptionally higher than its mean. Here, we defined the threshold to be two standard deviations of the number of messages per day, which is a common approach to detect outliers. For each pair of bars, the left bar (blue) shows the behavior for the pre-pandemic period while the right one (yellow) shows the behavior during the pandemic period. We performed a Kolmogorov-Smirnov test on the samples of both periods. The null hypothesis that both periods are from the same distribution is rejected with a p-value < 0.001. In the first period, 7.6% of the chats never exceed the threshold and therefore have a very regular conversation pattern. Furthermore, almost no chat has exceptional high numbers of messages for more than 10% of days. This is expected, as the threshold aims to detect outliers who by definition do not occur too frequently. The highest probability to pass two standard deviations is for 3–5% of the days, which corresponds to 1–2 outliers per month. For the pandemic year the probabilities for these shares decrease and instead the probability increases at the edge cases. The number of chats which never produce outliers with their messaging behavior is more than double at 15%. As such, these chats have a very constant role in the messaging activity of their users. This effect shows that during
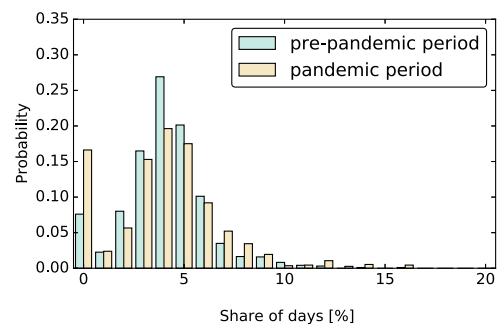

Fig. 7 Share of media for each 1-year period. The figure shows the share of sent media messages (e.g., images, video, and gifs) in the pre-pandemic year (left) compared to the pandemic year (right).

Fig. 8 Share of days with exceptionally high number of messages than mean. The figure shows the share of days for each chat where the number of messages was exceptionally higher (two standard deviations of the number of messages per day) than its mean. The values of the pre-pandemic period are colored in blue while the values of the pandemic period are marked in brown.

the pandemic year and the need of physical distancing, communication in WhatsApp became a more regular part of people's overall communication behavior. We speculate that regular in-person communication is now more often complemented by mobile communication. On the other hand, looking at chats having a share of 7% of days and more with an exceptionally higher number of messages than mean, an increase for the pandemic year is visible. The share of days for 7% to 10% equals 25 to 37 days of the 1-year periods. The main reason for this change is the sudden spike in messages during the first lockdown and is shown in Fig. 2. Furthermore, this trend is not visible when excluding the Mar/Apr periods from the full-year comparison. On the contrary, for the time after the lockdown, analysis indicates that more chats developed a regular communication pattern.

To sum up, concerning *RQ2* our evaluation shows that some changes in communication behavior were also visible beyond the lockdown, although other metrics have settled back to the initial value of the pre-pandemic period. In particular, we saw during the Mar/April 2020 lockdown that the media usage increased and changed its distribution. However, this cannot be said for the rest of the pandemic year, where media usage almost returned to the pre-pandemic level. If we look at the activity per time of day, however, we see that user behavior is still influenced by the pandemic and that, in contrast to the pre-pandemic period, midday continues to be the main communication time. By comparing the share of days with exceptionally higher than mean number of messages of the pre-pandemic and the pandemic year, the main observation is that 16.5% of the chats show a very regular conversation pattern having a 0% share of days with exceptionally higher than mean number of messages, which is almost twice as many as in the pre-pandemic year.

Conclusion
This work investigated the impact of the global COVID-19 pandemic on the communication behavior in the mobile instant messaging (MIM) app WhatsApp. In particular, we collected and evaluated a large dataset of 2577 private chat histories containing more than 25 million messages from almost 52,000 users. Based on this dataset, we investigated the short-term effects of the global COVID-19 lockdowns in March and April 2020 on the communication behavior compared to the same time span in the previous year 2019 and show that the communication behavior in WhatsApp significantly changed during the COVID-19 lockdown. We found that message frequency severely increased in the first lockdown months, reaching up to 2.58 times the average number of messages per day compared to the previous period. Moreover, the communication activity shifted during the day towards more communication at mid-day between 9 am and 2 pm, surpassing the previous activity peak in the early evening hours. In addition, more messages were sent after midnight from 12 am to 3 am compared to the pre-pandemic period, and the overall distribution of message types changed.

Furthermore, we analyzed the long-term effects of the ongoing COVID-19 pandemic by comparing the communication in the entire first pandemic year from March 2020 to February 2021 with the same time span in the pre-pandemic year, i.e., March 2019 to February 2020. Using our data set, we were able to confirm that there are both short-term and lasting effects of the pandemic on communication behavior in WhatsApp. A short-term effect could be observed for the number of sent messages, which returned to the pre-pandemic level after the initial spike in communication behavior in March and April 2020. In contrast, a persisting change was visible in the WhatsApp usage time over the course of the day. Finally, we noticed a shift towards more

regular communication. This means that the number of days with exceptionally high numbers of messages drastically reduced in the pandemic year. Thus, our results showed that the global COVID-19 lockdown and the ongoing pandemic severely altered the communication in WhatsApp chats compared to pre-pandemic time periods. This is in line with related works, which documented the impact of COVID-19 pandemic on many different areas of life. Nevertheless, in future work, it has to be checked for the long-term effects, which of them persist and which aspects return to pre-pandemic levels, not only for communication in MIM chats, but for all affected areas of life.

Data availability
The data set analyzed during the current study is not publicly available due to ongoing research in another project, but is available after project end from the corresponding author on reasonable request.

References
Amosun TS, Jianxun C, Rufai OH, Muhideen S, Shahani R, Shah Z, Koroma J (2021) WeChat usage during Covid-19 pandemic lockdown: the mediating role of online self-disclosure on quality of friendship and well-being. Global Knowledge, Memory and Communication, Vol. 71 No. 3, pp. 121–139.

Boettger T, Ibrahim IG, Vallis B (2020) How the internet reacted to Covid-19: a perspective from Facebook's edge network. In: Proceedings of the ACM Internet Measurement Conference, pp. 34–41

Casagrande M, Forte G, Tambelli R, Francesca F (2021) The coronavirus pandemic: a possible model of the direct and indirect impact of the pandemic on sleep quality in Italians. Nat Sci Sleep 13:191

Church K, de Oliveira R (2013) What's up with WhatsApp? Comparing mobile instant messaging behaviors with traditional SMS. In: Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services (MOBILE HCI), Munich, Germany

Dunford D, Dale B, Stylianou N, Lowther E, Ahmed M, de la Torre Arenas I (2020) Coronavirus: the world in lockdown in maps and charts. BBC News. https://www.bbc.com/news/world-52103747

Feldmann A, Gasser O, Lichtblau F, Pujol E, Poese I, Dietzel C, Wagner D, Wichtlhuber M, Tapiador J, Vallina-Rodriguez N et al. (2021) Implications of the Covid-19 pandemic on the internet traffic. In: Broadband Coverage in Germany; 15th ITG-Symposium. VDE. pp. 1–5

Garimella K, Tyson G (2018) WhatApp doc? A first look at WhatsApp public group data. In: Proceedings of the 12th international AAAI conference on Web and Social Media (ICWSM 2018), AAAI Press. pp. 511–517

Knop K, Öncü JS, Penzel J, Abele TS, Brunner T, Vorderer P, Wessler H (2016) Offline time is quality time. Comparing within-group self-disclosure in mobile messaging applications and face-to-face interactions. Comput Hum Behav 55:1076–1084

Li J, Zhou L, Heijden BVD, Li S, Tao H, Guo Z (2021) Social isolation, loneliness and well-being: the impact of wechat use intensity during the Covid-19 pandemic in China. Front Psychol 12:707667

Ling R, Chih-Hui L (2016) Microcoordination 2.0: social coordination in the age of smartphones and messaging apps. J Commun 66(5):834–856

Lutu A, Perino D, Bagnulo M, Frias-Martinez E, Khangosstar J (2020) A characterization of the Covid-19 pandemic impact on a mobile network operator traffic. In: Proceedings of the ACM Internet Measurement Conference, ACM. pp. 19–33

Mawarni ITA, Ratnasari N, Handayani AN, Muladi M, Aji Wibowo EP, Untari RS (2020) Effectiveness of WhatsApp in improving student learning interests during the Covid-19 pandemic. In 2020 4th International Conference on Vocational Education and Training (ICOVET), IEEE. pp. 248–252

Montag C, Błaszkiewicz K, Sariyska R, Lachmann B, Andone I, Trendafilov B, Eibes M, Markowetz A (2015) Smartphone usage in the 21st century: who is active on WhatsApp? BMC Res Notes 8(1):331

Nabity-Grover T, Cheung CMK, Thatcher JB (2020) Inside out and outside in: How the Covid-19 pandemic affects self-disclosure on social media. Int J Inform Manag 55:102188

Ohme J, Vanden Abeele MMP, Van Gaeveren K, Durnez W, De Marez L (2020) Staying informed and bridging "Social Distance": smartphone news use and

mobile messaging behaviors of flemish adults during the first weeks of the Covid-19 Pandemic. Socius 6:2378023120950190

Pinchas (2018) Why is WhatsApp more popular than SMS in Europe? December (2018) https://www.telemessage.com/why-is-WhatsApp-more-popular-than-sms-in-europe-infographic/

Pinto J, van Zeller M, Amorim P, Pimentel A, Dantas Patrícia, Eusébio E, Neves A, Pipa J, Santa Clara E, Santiago T et al. (2020) Sleep quality in times of Covid-19 pandemic. Sleep Med 74:81–85

Sandvine (2020) The global internet phenomena report February 2020, February 2020a. https://www.sandvine.com/hubfs/Sandvine_Redesign_2019/Downloads/2020/Phenomena/Mobile

Sandvine (2020) The global internet phenomena report: COVID-19 Spotlight May 2020, May 2020b. https://www.sandvine.com/hubfs/Sandvine_Redesign_2019/Downloads/2020/Phenomena/COVID

Schwind A, Seufert M (2018) WhatsAnalyzer: a tool for collecting and analyzing WhatAapp mobile messaging communication data. In: Proceedings of the 30th International Teletraffic Congress (ITC 30), IEEE.

Setiawan B, Iasha V (2020) Covid-19 pandemic: the influence of full-online learning for elementary school in rural areas. JPsd (Jurnal Pendidikan Sekolah Dasar) 6(2):114–123

Seufert M, Hoßfeld T, Schwind A, Burger V, Tran-Gia P (2016) Group-based communication in WhatsApp. In: IFIP Networking Conference (IFIP Networking) and Workshops. IEEE. pp. 536–541

Seufert M, Schwind A, Hoßfeld T, Tran-Gia P (2015) Analysis of group-based communication in WhatsApp. In: International Conference on Mobile Networks and Management. Springer. pp. 225–238

Statista Digital Market Outlook (2020) Digital economy compass 2020, November 2020. https://www.statista.com/study/83121/digital-economy-compass/#professional

Susilawati S, Supriyatno T (2020) Online Learning through WhatsApp group in improving learning motivation in the era and post pandemic Covid-19. J Pendidik: Teori, Penelit dan Pengembangan 5(6):852–859

Tahir Mehmood K (2021) Use of social media and WhatsApp to conduct teaching activities during the Covid-19 lockdown in Pakistan. Int J Pharm Pract 29(1):90–90

Tankovska H (2021) Most popular global mobile messenger apps as of January 2021, based on number of monthly active users, January 2021. https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/

Trevisan M, Vassio L, Giordano D (2021) Debate on online social networks at the time of Covid-19: an Italian case study. Online Soc Netw Media 23:100136

World Health Organization (WHO) (2020) Archived: WHO Timeline—COVID-19. https://www.who.int/news/item/27-04-2020-who-timeline—covid-19

Xia Y, Zhang Y, Zhong Z, Yan G, Lim C, Singh Ahuja S, Bali S, Nikolaidis A, Ghobadi K, Ghobadi M (2021) A social network under social distancing: risk-driven backbone management during Covid-19 and beyond. In: NSDI. pp. 217–231

Competing interests

The authors declare no competing interests.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Informed consent

Every user was informed via the Terms of Usage and Privacy Policy of the implemented tool that they share personal data with us, which we only stored and analyzed in an anonymized format. Moreover, by using the tool, users consented to participate in a scientific study and agreed that the collected data is evaluated and the study results are published.

Additional information

Correspondence and requests for materials should be addressed to Anika Seufert.

Reprints and permission information is available at http://www.nature.com/reprints

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