

WhatsApp, Doc?

A First Look at WhatsApp Public Group Data*

Kiran Garimella

EPFL, Switzerland

kiran.garimella@epfl.ch

Gareth Tyson

Queen Mary University, London

gareth.tyson@qmul.ac.uk

Abstract

In this dataset paper we describe our work on the collection and analysis of public WhatsApp group data. Our primary goal is to explore the feasibility of collecting and using WhatsApp data for social science research. We therefore present a generalisable data collection methodology, and a publicly available dataset for use by other researchers. To provide context, we perform statistical exploration to allow researchers to understand what public WhatsApp group data can be collected and how this data can be used. Given the widespread use of WhatsApp, our techniques to obtain public data and potential applications are important for the community.

1 Introduction

The Short Message Service (SMS) was initially envisaged as a feature of the GSM standard. It enabled mobile devices to exchange short messages of up to 160 characters. Despite its auxiliary nature, it rapidly became popular; in 2010, 6.1 trillion SMS were sent (ITU 2010). However, this is beginning to be surpassed by the emergence of several Internet-based messaging apps, *e.g.*, WeChat, Telegram and Viber. Although these apps have pockets of dominance, the clear market leader is **WhatsApp** (Daniel Sevitt 2016). For example, in India, over 94% of all Android devices have the app installed with an average of 78% of current installs using it daily.

The reasons for its dominance are numerous. Released in 2009, WhatsApp was the forerunner of mobile messaging apps. At this time, many mobile subscribers were charged for sending SMS — WhatsApp offered a free equivalent, whilst allowing users to maintain many of the convenient aspects of SMS, *e.g.*, identification via phone numbers. WhatsApp also introduced powerful new features, such as the ability to include multimedia content and create shared groups. In 2017, WhatsApp reached 1 billion users each day, with 55 billion daily messages being sent (Deahl 2017).

This suggests that a major portion of online interactions take place via WhatsApp. Indeed, its popularity far exceeds more traditional messaging services like Skype (Daniel Sevitt 2016). However, its group functionality and easy in-

tegration of multimedia content indicates that usage may differ significantly from these other platforms, particularly SMS. This is confirmed in social studies that have found that WhatsApp tends to be used in a more conversational and informal manner amongst close social circles (Church and de Oliveira 2013). A particularly novel aspect of WhatsApp messaging is its close integration with public groups. These are openly accessible groups, frequently publicised on well known websites,¹ and typically themed around particular topics, like politics, football, music, *etc.* This constitutes a radical shift from the bilateral nature of SMS. As such, we argue that these public WhatsApp groups warrant study in their own right. More generally, although past studies have investigated WhatsApp usage via methodologies such as interviews (Church and de Oliveira 2013), we believe it is important to perform both large-scale and data-driven analyses of its usage.

With this in mind, this **dataset paper** presents a methodology to collect large-scale data from WhatsApp public groups. To demonstrate this, we have scraped 178 public groups containing around 45K users and 454K messages. Such datasets allow researchers to ask questions like (i) Are WhatsApp groups a broadcast, multicast or unicast medium? (ii) How interactive are users, and how do these interactions emerge over time? (iii) What geographical span do WhatsApp groups have, and how does geographical placement impact interaction dynamics? (iv) What role does multimedia content play in WhatsApp groups, and how do users form interaction around multimedia content? (v) What is the potential of WhatsApp data in answering further social science questions, particularly in relation to bias and representability?

We begin by presenting related studies that have either focused on WhatsApp or messaging services more generally (§2). Due to the difficulty in data collection, most of these studies rely on qualitative methods and interviews/surveys. Our dataset therefore constitutes the first large-scale public WhatsApp data source. We then describe our data collection methodology, which involves scraping a list of public WhatsApp groups, subscribing to them, and then monitoring them such that all communications can be imported into an easy-to-use schema (§3). With this data, we then proceed to per-

*Code and dataset from this paper can be found at <https://github.com/gvrkiran/whatsapp-public-groups>
Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

¹For example, <https://joinwhatstappgroup.com/>

form a basic characterisation, outlining its key trends (§4). We particularly focus on exploring the potential, as well as the biases we see in the dataset. We conclude that collecting large-scale public messaging data with WhatsApp is feasible, and one can obtain a broad geographical coverage (§5). However, we also find that diversity amongst groups is high (both in terms of activity levels, geography and topics covered). Hence, a careful selection of seed groups is paramount for meaningful results. In summary:

- We show the possibility of collecting publicly available WhatsApp data.
- Using the above approach, we collect an example dataset of 178 groups, containing 45K users and 454K messages.
- We characterise the patterns of communication in these groups, focussing on the frequency, types and topics of messages.
- We show the applicability of such data in answering new social science research questions.
- We release an anonymised version of the data and all the code used to allow others to collect targeted datasets on groups relevant to their research.

2 Related work

We see two major themes of related work: (*i*) studies that have explored social communication patterns on SMS and similar messaging services; and (*ii*) studies that have focused on WhatsApp itself.

Studies of Messaging There have been a large number of studies exploring user behaviour regarding messaging. Due to popularity amongst teenagers, many studies have focused on their usage patterns. This has included work across various countries, including Finland (Kasesniemi and Rautainen 2002), Norway (Ling and Yttri 2002), the United Kingdom (Grinter and Eldridge 2001; 2003; Faulkner and Culwin 2004) and the United States (Battestini, Setlur, and Sohn 2010). Generally, services like SMS have been found to be primarily used within close social groups for activities such as general conversation, planning and coordination (Grinter and Eldridge 2001). This is driven by its low cost, ease of use and lightweight nature. Other research has focused on the language used, including the emergence of text-based slang (Grinter and Eldridge 2003) and usage of messaging across different age ranges (Kim et al. 2007). A key limitation of these studies has been the focus on qualitative methodologies, *e.g.*, interviews, surveys, focus groups. One study collected quantitative data via the installation of a logging tool on user devices (Battestini, Setlur, and Sohn 2010). By recruiting 70 participants, they analysed 58K sent messages. Although powerful, this approach is largely non-scalable and creates datasets that are challenging for public use due to privacy constraints. Other messaging apps, such as WeChat (Huang et al. 2015), have been explored at scale although the focus has not been on the content and interactions. Instead, coarser analyses have been performed, *e.g.*, size of messages. Studies that have explored more social features have, again, limited themselves to small-scale

surveys (Lien and Cao 2014). It is worth noting that there have also been several studies exploring messaging patterns within other community mediums, *e.g.*, Reddit (Singer et al. 2014), 4chan (Bernstein et al. 2011; Hine et al. 2017) and IRC (Rintel, Mulholland, and Pittam 2001). We consider such platforms orthogonal to WhatsApp, and therefore do not focus on them here.

Studies of WhatsApp There have been a small number of studies that have inspected the usage of WhatsApp specifically. Due to the differences between WhatsApp and SMS, these deserve discussion in their own right. These studies tend to centre on WhatsApp usage within given settings. For example, there have been studies inspecting how students and teachers interact via WhatsApp (Bouhnik and Deshen 2014), as well as the impact WhatsApp usage may have on school performance (Yeboah and Ewur 2014). Similar studies have been performed within medical settings to understand how WhatsApp facilitates communication amongst surgeons (Wani et al. 2013; Johnston et al. 2015). The common limitation of these studies is their reliance on small populations and qualitative methodologies (*e.g.*, interviews). Although important, this provides little insight into more general purpose usage across “typical” users. Church *et al.* also performed a direct comparison of SMS *vs.* WhatsApp, finding that interviewees used WhatsApp more often, confirming its growing importance (Church and de Oliveira 2013).

In contrast to the above studies, which rely on surveys and interviews, (Rosenfeld et al. 2016) took a quantitative approach by harvesting WhatsApp data directly from 92 volunteers. Due to the private nature of the messages, the authors focused on metadata rather than message content, *e.g.*, length of text. Montag *et al.* took a similar approach, asking 2418 users to download an app that records usage (Montag et al. 2015). Both works are highly complimentary to our own; the main difference is that we focus on *public* rather than private WhatsApp communications, allowing us to yield datasets with orders of magnitude more users. This is because the intrusive nature of the data collection in these other studies makes it difficult to scale-up beyond small numbers of users.

3 Data collection

This section delineates the data collection methodology, as well as its limitations and ethical considerations. Both the tools and datasets are publicly available.

3.1 Data Collection Methodology

We begin by detailing our data collection methodology. We intend this to be generalisable across any set of WhatsApp groups or, indeed, other online messaging services that support public groups. For this, we only required a single low-capacity compute server, alongside a working mobile device with WhatsApp installed. A single working phone number is required, such that the WhatsApp SMS confirmation can be received to register the device. Once these tools are in place, the data collection contains two steps.

Step 1 First, it is necessary to acquire a set of pub-

lic groups for data collection. We are not prescriptive in how these are obtained. For example, some researchers may wish to manually curate a list or target just a small number of highly specific groups. This is supported by a number of existing websites that index public groups (*e.g.*, joinwhatappgroup.com/). We, however, took a more large-scale approach. We used the Google search engine, and other focussed websites, to compile a list of public groups. This was attained by searching for links that contain the suffix of `chat.whatsapp.com`.² This gave us a list of 2,500 groups.

Next, we randomly sampled 200 groups from this list and joined them using an automated script. The script uses a browser automation tool, Selenium and the `web.whatsapp.com` web interface to automate the joining process. Note that the web interface needs a single time sign in (via scanning a QR code) with the same account as the Android device we will use to subsequently collect the data. At the conclusion of Step 1, we had a dedicated WhatsApp account subscribed to the full set of groups with little human intervention in the process. Hence, this can easily scale to much larger sets of groups.

Step 2 Once we joined the groups, we started to receive updates on the phone. As WhatsApp implements end-to-end encryption³ it is naturally difficult to passively collect data on the device (*e.g.*, via Wireshark). Fortunately, WhatsApp stores all messages received within a simple sqlite database on the local device. This made it trivial to extract the data being collected periodically from the device (once the storage began to fill). To make this feasible, however, it was necessary to use the encryption keys to decrypt the stored version of the messages.⁴ We therefore used the technique of Gudipaty *et al.* (Gudipaty and Jhala 2015) to extract the storage key and decrypt all messages.⁵ Overall, we collected data for 178 groups,⁶ containing 45,794 users, and 454,000 messages over a 6 month period (May-Oct 2017).

We will share the code and (anonymised) data after the paper is accepted.

3.2 Ethical Considerations

Clearly the above methodology has the capacity to collect large bodies of data containing messages sent by individuals from around the world. There are therefore certain privacy considerations that must be taken into account. Most notably, individual phone numbers should *not* be collected and/or released. To anonymise users, we allocate each phone number a unique identifier after extracting the appropriate

²An example of a public WhatsApp group: <https://chat.whatsapp.com/BZp0Ye2eoRp2TWhQe7ixvO>

³<https://www.whatsapp.com/security/>

⁴Messages are both transmitted and stored in an encrypted form

⁵The encryption key can also be obtained in a much simpler manner with a rooted Android phone, *e.g.* see <http://jameelnabbo.com/breaking-whatsapp-encryption-exploit/>.

⁶22 out of the 200 groups were either removed or had no activity.

country code. We also advice researchers to delete the WhatsApp device database after data has been extracted from the device (because the WhatsApp database will continue to store the phone number). To further guarantee privacy, we also do not release message content in our public dataset (just metadata).

Researchers should also be careful regarding which types of groups they choose to scrape. Although all groups are public and therefore users are aware that their messages will be seen by unknown parties, it is worth noting that there are a wide diversity of group types. These include those of an adult nature, which some researchers may wish to avoid, *cf.* (Tyson et al. 2015) for further discussion. Moreover, researchers will have no control over the content sent via the groups; hence, there is a risk of receiving unsavoury or even illegal multimedia content. Our advice is therefore to disable the automatic downloading feature on the device running WhatsApp (this is also helpful for improving scalability).

Finally, we emphasise that the privacy policy for WhatsApp groups states that a user shares their messages and profile information (including phone number) with other members of the group (both for public and private groups).⁷ Group members can also save and email upto 10,000 messages to anyone.⁸ Our paper provides automated tools for this process.

4 Characterising WhatsApp groups

To provide context for the applicability of WhatsApp group data, we next characterise its basic properties. We particularly focus on identifying the issues and biases that may occur within such data. Although we utilise our collected dataset to underpin this, other researchers can apply a similar methodology to acquire data in their target domains.

4.1 How much data can be collected?

Over the 6 month period, we collected data from 178 groups. Each group had an average of 143.3 participants (median 127), with the largest group observed containing 314 participants.⁹ In total, 454K messages were collected, spanning 45K users. Figure 1 presents the number of messages sent per-user. Unsurprisingly, the distribution is highly skewed with the top 1% of users generating 37% of all messages. Around 10K users (25%) have more than 5 messages. The remaining 75% of the users are mostly consumers of information.

Figure 2 shows how these messages are distributed across groups. We find that over 30% of the groups have under 1000 messages during the 6 month measurement period. Despite this, there are a small number of highly active groups — the most active generated 11K messages overall. This indicates there is a high degree of scope for optimisation with researchers being able to get significant volumes of messages

⁷<https://www.whatsapp.com/legal/>

⁸<https://faq.whatsapp.com/en/android/23756533/>

⁹Note, at the time of writing there is a default maximum of 256 group members per group, which can be increased manually.

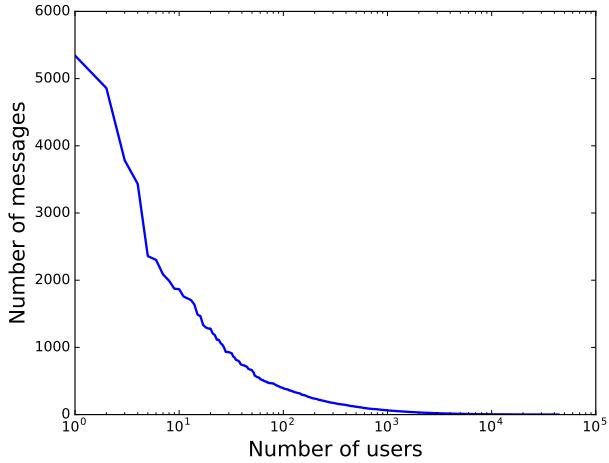


Figure 1: Activity of users in our dataset. 75% of the users have less than 5 messages.

from just a few groups. Data from the top 10 groups would yield in excess of 80K messages (18% of our overall set). As such, it is clear that WhatsApp can be effectively used for garnering significant social datasets.

4.2 Where are users located?

The above has shown that large quantities of social data *can* be collected from WhatsApp groups. We next ask what geographical biases may be contained within such data. Each user is associated with a phone number. By examining the country code, it is possible to geolocate users based on their registered country. This has the benefit of not changing whilst users are visiting other countries (unlike datasets based on GPS or IP geolocation).

Figure 3 presents a heatmap of user locations. The top countries include India (25K), Pakistan (3.6K), Russia (3K), Brazil (2K) and Colombia (1K). This immediately confirms a significant geographical bias, although not towards the United States as one would typically expect. This may therefore be considered as a positive point by many social science researchers. For example, we see many users in developing regions, *e.g.*, in Africa, Nigeria has 959 users, whilst in South America, Colombia has 1,073 users. Hence, we posit that these datasets may offer effective cultural vantage into developing regions as well as developed ones.

This diversity is also mirrored in the make-up of individual groups. Remarkably, we do not find *any* groups that are limited to a single country. Instead, all groups contain members from multiple countries. Figure 4 presents a histogram of the number of countries contained within each group. It can be seen that significant international communities are present within the groups. 85% of groups have members from over 10 different countries. Again, this indicates that the data offers a vantage into globalised communities that easily cross national boundaries. We looked at the 5 groups that have users from more than 30 countries, to

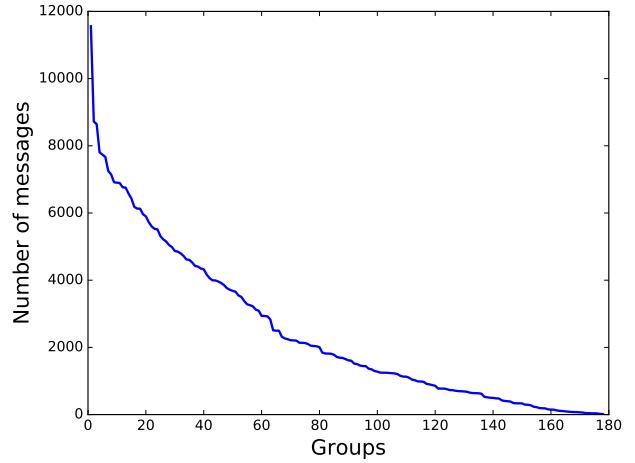


Figure 2: Number of messages per group. Over 30% of the groups have less than a 1000 messages in 6 months.

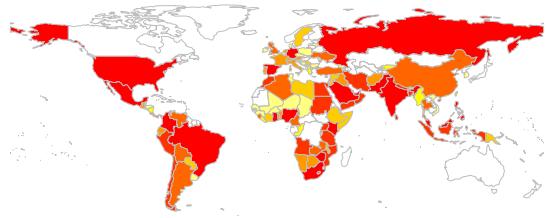


Figure 3: Location of users in our dataset. Brighter shades of red indicate higher number of users.

find that they varied in type, including sex, English learning, YouTube videos, *etc.*

Another property of geography is language. We automatically inferred the language of a message using Lui *et al.* (Lui and Baldwin 2011). Note that our analysis on language depends on the performance of their model. Across the 178 groups, we observe 59 languages which have at least 200 messages sent. Table 1 presents a breakdown of the most popular languages. Unsurprisingly, English is most prominent with in excess of 137K messages. This is followed by Hindi, and other Indian languages such as Gujarati, Tamil and Marathi. Although a powerful feature in itself, this does significantly complicate analysis. Unfortunately, many groups contain messages of multiple languages, making deeper social analysis even more challenging. This is not just occasional messages as we find that 33% of groups have less than 50% of messages in a single language.

4.3 What is sent?

We now progress to explore the content of what is sent within the groups. We remind the reader that this is heavily impacted by the choice of groups being scraped. As previously stated, we collected 454K messages overall. From

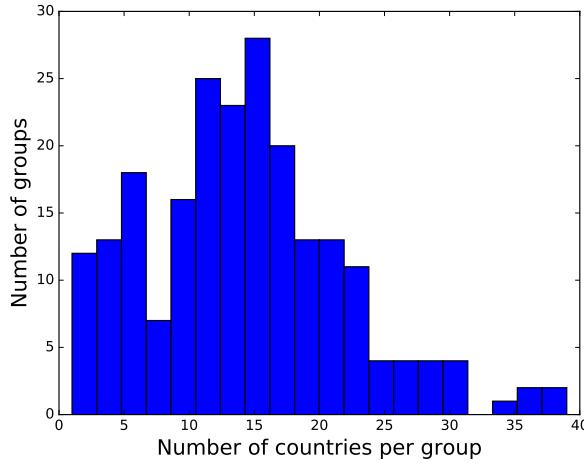


Figure 4: Histogram showing number of countries users in a group belong to. A majority of the groups have users from more than 10 countries.

# Messages	Language
137527	English
78333	Hindi
13063	Spanish
7525	Gujarati
5341	Tamil
5123	Chinese
4193	Marathi
2942	German
2930	Polish
2349	Italian

Table 1: Top 10 most popular languages as measured by number of messages sent.

these, 9.1% were images, 3.6% were videos, and 0.7% audio; the rest were text. The average image size is 101KB, whilst the average video is a non-negligible 4.6MB. The average length of the text messages 582 characters (median 136 characters).

As well as content, we observe a large number of URLs being shared — a remarkable 39% of messages contain web links. This offers a powerful tool for researchers wishing to explore social web content popularity. Table 2 presents the most popular domains shared via WhatsApp, as well as their Alexa Ranking. Although we observe many of the international hypergiants (*e.g.*, Google, YouTube) we also observe a wide range of fringe websites. There is little correlation between the popularity of the domain in our WhatsApp data and its popularity on Alexa. Of course, this is partly driven by the geographical distribution of the user base; for example, `lootdealsindia.in` has a global rank of 917,011 but an Indian ranking of 83,911. Despite this, it is clear that WhatsApp groups may offer an effective vantage into lesser known web content and how it is accessed by fringe communities.

# Messages	Domain	Alexa Rank
59883	youtube.com	2
37270	whatsapp.net	614,880
12239	amazon.in	90
7141	google.com	1
5395	whatsapp.com	69
3979	blogspot.com	63
1989	wowapp.com	78,514
1218	flipkart.com	161
1144	lootdealsindia.in	917,011
1032	marugujraat.com	6,217,479
952	kamalking.in	799,769
630	dealvidhi.com	2,895,020
455	facebook.com	3
453	mydealone.com	7,882,171
431	msparmar.in	5,008,742
405	newsdogshare.com	163,914
402	newsdesire.com	N/A
346	sex.xxx	N/A
324	ojasinfo.com	2,949,092
323	jobdashboard.in	324,811

Table 2: Most popular domains within URLs shared via WhatsApp groups. `whatsapp.net` urls mostly contain multimedia. `google.com` is mostly for sharing playstore apps (`play.google.com`).

We can also inspect the temporal trends of when these messages are sent. Figure 5 depict the total number of messages sent on each day of the week for the top 20 groups in terms of activity. Two noteworthy things can be observed. First, the greatest activity occurs on weekdays, rather than weekends. Second, the peak day for most groups is Wednesday. Why this might be is unclear, however, it is evident that this holds across many groups. 79% of all the 178 groups peak on a Wednesday. This trend is in line with other social networks like Facebook and Twitter, where previous studies have revealed increased activity during weekdays with peaks on Wednesday.¹⁰ It is also worth briefly noting that very few (under 2%) of these messages are replies.¹¹ This is a feature that is rarely used, therefore making it difficult for researchers to formally understand who is talking to whom within groups.

4.4 What topics are captured?

Finally, we inspect the topics captured within the groups. There is no formal taxonomy of topics within WhatsApp and, thus, it is necessary for researchers to manually inspect and classify the groups under study. We manually annotated the 178 groups we collected into a set of categories. From our WhatsApp dataset, we find several types of groups with significant followings: (i) *generic groups* – ‘funzone’, ‘funny’, ‘love vs. life’, *etc.* (70 groups); (ii) *adult groups* – ‘XXX’, ‘nude’, *etc.* (19 groups); (iii) *political aligned*

¹⁰<http://bitly.tumblr.com/post/22663850994/time-is-on-your-side>

¹¹Users can directly send replies to other messages

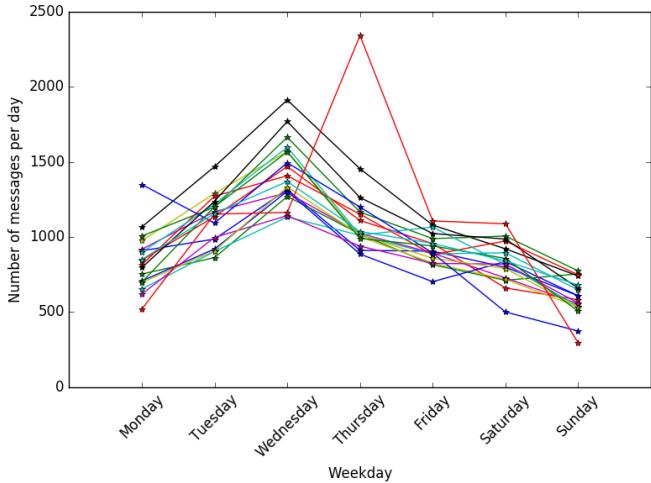


Figure 5: Number of messages sent per day for the top 20 groups with highest activity.



Figure 6: Word cloud generated from group titles. All 2500 groups identified in Step 1 of the methodology were used.

groups – mostly Indian political parties (15 groups); (iv) *movies/media* — ‘box office movies’, fan groups, anime, etc (17 groups); (v) *spam* — deals, tricks (14 groups); (vi) *sports* — football ('football room'), cricket ('world cricket fans'), etc. (12 groups); (vii) other – job posts, education discussion, tech, activism, etc. (23 groups);

Hence, researchers wishing to focus on any of these topics could certainly do so via WhatsApp data. The largest group is “DISFRUTA AL MAXIMO” (enjoy to the fullest) which contains 11K messages, primarily based in Colombia, followed by “No life without cricket” (8.7K messages, India), and “Football room” (7.7K messages, Nigeria). Again, we emphasise that these statistics are biased by our choice of groups, however, their diversity confirms that it would be possible for many different topics to be explored via these groups. Briefly, to provide finer-grained vantage of the topics discussed, we can inspect the words used within the group titles. Figure 6 presents a word cloud generated using the group titles. In-line with the above topics, we observe regular discussions related to concepts such as nudity, videos and cash, as well as geographical indicators such as India.

5 Conclusion & Discussion

The paper has provided tools to collect WhatsApp data for the first time. The dataset we collected is a random sample of 178 public groups, however, the principle behind this paper is to show that large scale data collection from WhatsApp groups is feasible. Such datasets, if collected with a predefined goal in mind, have immense consequences and open up new areas of research.

As well as presenting our methodology, we have also performed a basic characterisation of our dataset to highlight its key features. This has revealed potential bias in factors such as geographical user distribution. However, rather than being a limitation, we believe such bias could be exploited. For example, one important finding is the ability to collect data both globally and across borders. Although this naturally covers highly connected regions such as Europe and North America, we also observe a significant number of users in developing regions. Thus, we argue that WhatsApp may be particularly useful for offering vantage into such regions (which are often overlooked in mainstream research). For example, in India alone, it is estimated that by 2020, 400 million new users who have never been a part of the digital data realm, will join the Internet. The popularity of WhatsApp means that it could act as a powerful research tool for understanding this growing use. With this in mind, we conclude by listing a few ambitious questions that we believe WhatsApp group data may be able to help answer:

1. Can we find the emergence of new social institutions from WhatsApp group data? Given this new ecosystem of connectivity that empowers users, new institutions such as markets (micro work, virtual trading), money (e.g., WeChat money, AliPay, PayTM), and social organisations (trade unions) may emerge. How would such trends be reflected in WhatsApp activity?
 2. Can we understand the role of these new institutions in shaping the economic, social and wellbeing of the people who constitute these institutions? For instance, understanding the effects of new markets on patterns of migration and assimilation between villages and cities. WhatsApp data could potentially expose these patterns as users come and go between groups, and as new groups emerge to reflect these institutions.
 3. Can we use this data to explore and understand how information such as “fake news” spreads through communities. This is particularly relevant as fake news is a significant issue on WhatsApp, especially in countries with low levels of digital literacy.¹² More generally, how does multimedia content propagate through (and spread between) such groups?
 4. Can we make use of the insights taken from WhatsApp groups to create algorithms to help deliver better services to users, which can improve their way of life? For example, (i) *Livelihood*: micro-matching jobs and talents, (ii) *Wellbeing*: using WhatsApp-shared image analysis for automated medical diagnoses, (iii) *Education*: Delivering

¹²<http://bit.ly/2DuStFn>

the right content to the right people — educating farmers with crop season information, etc. Each of these topics could benefit from their implementation over WhatsApp, e.g., using groups to share relevant employment information in communities.

The above topics go well beyond the scope of this initial work. However, as a popular medium for communication in many parts of the world, we argue that WhatsApp should be given equal attention to that of other social media services, e.g., Twitter. We hope that this work, and its associated tools, can act as a platform for other research to build atop of.

References

- Battestini, A.; Setlur, V.; and Sohn, T. 2010. A large scale study of text-messaging use. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, 229–238. ACM.
- Bernstein, M. S.; Monroy-Hernández, A.; Harry, D.; André, P.; Panovich, K.; and Vargas, G. G. 2011. 4chan and /b/: An analysis of anonymity and ephemerality in a large online community. In *ICWSM*, 50–57.
- Bouhnik, D., and Deshen, M. 2014. Whatsapp goes to school: Mobile instant messaging between teachers and students. *Journal of Information Technology Education: Research* 13:217–231.
- Church, K., and 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*, 352–361. ACM.
- Daniel Sevitt. 2016. Popular messaging apps by country. <https://www.similarweb.com/blog/popular-messaging-apps-by-country>.
- Deahl, D. 2017. More than 1 billion people are now using whatsapp every day. <https://www.theverge.com/2017/7/27/16050220/whatsapp-1-billion-daily-users-250-million-whatsapp-status>.
- Faulkner, X., and Culwin, F. 2004. When fingers do the talking: a study of text messaging. *Interacting with computers* 17(2):167–185.
- Grinter, R. E., and Eldridge, M. A. 2001. y do tngrs luv 2 txt msg? In *ECSCW 2001*, 219–238. Springer.
- Grinter, R., and Eldridge, M. 2003. Wan2tlk?: everyday text messaging. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 441–448. ACM.
- Gudipaty, L., and Jhala, K. 2015. Whatsapp forensics: decryption of encrypted whatsapp databases on non rooted android devices. *Journal of Information Technology & Software Engineering* 5(2):1.
- Hine, G. E.; Onaolapo, J.; De Cristofaro, E.; Kourtellis, N.; Leontiadis, I.; Samaras, R.; Stringhini, G.; and Blackburn, J. 2017. Kek, cucks, and god emperor trump: A measurement study of 4chan's politically incorrect forum and its effects on the web. In *ICWSM*, 92–101.
- Huang, Q.; Lee, P. P.; He, C.; Qian, J.; and He, C. 2015. Fine-grained dissection of wechat in cellular networks. In *Quality of Service (IWQoS), 2015 IEEE 23rd International Symposium on*, 309–318. IEEE.
- ITU. 2010. The world in 2010: The rise of 3g. <http://www.itu.int/ITU-D/ict/material/FactsFigures2010.pdf>.
- Johnston, M. J.; King, D.; Arora, S.; Behar, N.; Athanasiou, T.; Sevdalis, N.; and Darzi, A. 2015. Smartphones let surgeons know whatsapp: an analysis of communication in emergency surgical teams. *The American Journal of Surgery* 209(1):45–51.
- Kasesniemi, E.-L., and Rautiainen, P. 2002. 11 mobile culture of children and teenagers in finland. *Perpetual contact* 170.
- Kim, H.; Kim, G. J.; Park, H. W., and Rice, R. E. 2007. Configurations of relationships in different media: Ftf, email, instant messenger, mobile phone, and sms. *Journal of Computer-Mediated Communication* 12(4):1183–1207.
- Lien, C. H., and Cao, Y. 2014. Examining wechat users motivations, trust, attitudes, and positive word-of-mouth: Evidence from china. *Computers in Human Behavior* 41:104–111.
- Ling, R., and Yttri, B. 2002. 10 hyper-coordination via mobile phones in norway. *Perpetual contact: Mobile communication, private talk, public performance* 139.
- Lui, M., and Baldwin, T. 2011. Cross-domain feature selection for language identification. In *In Proceedings of 5th International Joint Conference on Natural Language Processing*. ACL.
- Montag, C.; Błaszkiewicz, K.; Sariyska, R.; Lachmann, B.; Andone, I.; Trendafilov, B.; Eibes, M.; and Markowetz, A. 2015. Smartphone usage in the 21st century: who is active on whatsapp? *BMC research notes* 8(1):331.
- Rintel, E. S.; Mulholland, J.; and Pittam, J. 2001. First things first: Internet relay chat openings. *Journal of Computer-Mediated Communication* 6(3):0–0.
- Rosenfeld, A.; Sina, S.; Sarne, D.; Avidov, O.; and Kraus, S. 2016. Whatsapp usage patterns and prediction models. *ICWSM/IUSSP Workshop on Social Media and Demographic Research*.
- Singer, P.; Flöck, F.; Meinhart, C.; Zeitfogel, E.; and Strohmaier, M. 2014. Evolution of reddit: from the front page of the internet to a self-referential community? In *Proceedings of the 23rd International Conference on World Wide Web*, 517–522. ACM.
- Tyson, G.; Elkhatab, Y.; Sastry, N.; Uhlig, S.; et al. 2015. Are people really social in porn 2.0? In *ICWSM*, 236–444.
- Wani, S. A.; Rabah, S. M.; AlFadil, S.; Dewanjee, N.; and Najmi, Y. 2013. Efficacy of communication amongst staff members at plastic and reconstructive surgery section using smartphone and mobile whatsapp. *Indian journal of plastic surgery: official publication of the Association of Plastic Surgeons of India* 46(3):502.
- Yeboah, J., and Ewur, G. D. 2014. The impact of whatsapp messenger usage on students performance in tertiary institutions in ghana. *Journal of Education and practice* 5(6):157–164.