An Empirical Study of Shift from SMS to Chat-Apps among University Student

Bhopendra Singh, Vinod Kumar Shukla, Sonia singh

Abstract: Due to the high penetration of smartphones and growing availability of mobile internet in India, there is a shift from SMS to Mobile Messaging Applications (Chat Apps). This is expected as Chat Apps provide several advantages over SMS as group chat, sharing of media, privacy, emotes or stickers, and several more features at almost no cost. This study attempts to further understand the shift from SMS to Chat Apps by analysing the mobile usage patterns of over four thousand Amity University students across India. Unlike previous studies where questionnaires and other self-reporting data collection tools are used, this study makes use of a Mobile Application, which automatically collects relevant information from the participant's smartphone. This eliminates misinformation or bias since the data is collected without any direct involvement of the participant except for his or her approval. The sample was significantly diverse with information from students under 95 institutions under 6 Amity Universities across 4 adjoining States of India. Using statistics and data mining techniques, six different domains were assessed: Demographics, Smartphone Devices, Network Service Providers, Internet Usage, SMS and Call Usage, and Mobile Application Usage. Not only is this paper relevant to Telecommunication Industry, but also to Mobile App Developers by determining the most popular app categories and the average amount of internet data consumed by students of Indian Universities.

Index Terms: Mobile Messaging Apps, Telecommunication Industry, OTT Apps, SMS, Call, Social Media

I. INTRODUCTION

The shift to Chat Apps is increasing day by day with the growth of online application and its popularity on social networking. Consumers are showing keen interest in new technology, therefore market is moving from home computer systems to mobile devices [1]. Chat App has proved to be a very disruptive technology [2]. Mobile Instant Messaging has (MIM) tools have become popular, particularly among young people. The Oxford Analytica Daily Brief, explained that a main reason behind the success of Chat Apps like WhatsApp was the extremely high reach of the mobile Internet [3]. In addition, Social networking has continued to evolve providing newer ways for people to interact with people, brands and events [4]. There are 8 distinct factors that led to the adoption of Chat Apps, namely, pricing, social influence, nature or intent, Community & sense of connection, Immediacy & Privacy Concerns, Reliability, Choice of Technology and Coping Mechanisms [5].

Chat Apps are growing very popular in India. There are number of reasons why this shift is expected. First, India is a very price sensitive market. This is why both BBM and WhatsApp have become quite popular. In addition, there has been increased smartphone penetration (dominated by low-end Chinese made smartphones). Both of these factors have led to rapid growth in OTT messaging [6]. Because of this, Telecom operators are at a crucial point in the progression of their industry. Their traditional methods of income (SMS & Calls) are failing and making it extremely difficult for them to build extremely fast mobile networks (which are in turn growing in demand). Unless these mobile carriers design better business models, they will be unable to survive in such a competitive ecosystem [7].

There are two distinct kinds of mobile phone usage studies that have been done in the past. One type deals with analysing the usage behaviour by using questionnaires, interviews and diary writing methods. These types of self-reported data are susceptible to a number of problems like personal prejudice or the inability to recollect something relevant and important to the study. The other type uses automatic collection and analysis of mobile phone logs. This type is often used for longer time durations and for a far greater number of participants [8][9][10][11].
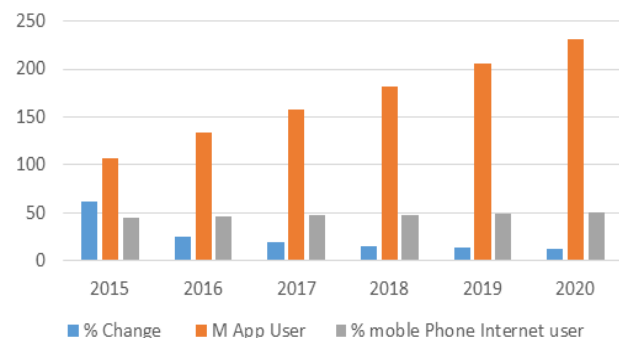

Fig.1. Mobile Messaging App uses

This study is designed to further understand the shift from SMS to Chat Apps by young smartphone users. This is done with consideration to Telecommunications Carriers by measuring all chargeable entities like sms sent, calls made and internet data used. One of the motivations of this study is also continuous growth of mobile app users and mobile phone users in India. This is represented in following Fig.1, which shows the growth from 2015 to 2020. As per the forecast done by eMarketer's approximately 25% of the world population will be using mobile messaging apps[12]

II. RESEARCH METHODOLOGY

Unlike the more conventional methods of data collection, which require the participant to undergo self-reporting activities like diary writing or questionnaires, this paper, requires the participants to install an Android Application, which collects relevant information. Factual data records have been collected automatically behind the scenes leveraging both the capabilities of a smart phone as well as utilizing the Profile Information accessible from the online University Student Portal account.

A. Objectives

- To assess the social activity of college students in terms of **Calls, SMS and net usage**.
- To assess the top categories of **apps installed** and rank them on the basis of count and network data used.
- To assess the top **mobile carriers and smartphones** used by university students.

B. Sample Size

Data from 5387 users has been collected for this paper. It should be noted that these are all Android Mobile users who are currently students of Amity Universities across India. However, only **4465 user data have been used for analysis** after removal of invalid or inconsistent information.

C. Methodology Used

This paper makes use of a **logging tool** to collect information about the participant without his direct involvement in data entry process. This logging tool specially created for this paper, collects certain information from the user leveraging the capabilities provided by the Android Operating System as well as the Profile Information provided in the participant's Online Student Portal.

The most significant advantage seen by using a data collection tool is that the data is significantly more accurate without being misrepresented or sabotaged by the participant during the feedback process. For example: In Questionnaires, the *Age* of the respondent is collected as part of the demographic variables. These questionnaires get filled over a period of time and the Age collected during the first few days of the data collection process may have changed towards the last few days of the same. In this logging tool, the *date of birth* has been collected from the participants from which their ages are computed with respect to a fixed date (which in this case is January 1st, 2015).

A total of 5387 user data has been collected for this paper. This was possible only by integrating the logging tool into a popular Android Application used by Amity University Students. *AMIZONER* is an Android Application freely available for download on the Google Play Store. It is an attendance planner that is popular among students of Amity University and currently has over 30,000 installations in total. The logging tool (that has been developed specifically for this paper) has been integrated to this application, AMIZONER.

Leveraging the large user base of AMIZONER, the logging tool will collect relevant information to this research and upload it to a server. This server will collect a massive amount of information from the mobile handsets of thousands of users. These thousands of unique records are then combined, aggregated, filtered, analysed, and reformatted using programs developed specifically for this paper.

The logging tool was installed with prior permission of the student. Before the user installed the AMIZONER app on his or her phone, he or she was explained that information would be gathered for this dissertation project with a message. He or she could choose to opt out. Also, various techniques have been used to protect the privacy of the users. For example, text-message content has not been collected. Instead, only the count of words, characters and emotes were collected.

D. Tools Used

A wide variety of tools have been used in this study. These are explained one by one segregated on the basis of the 4 stages involved [11].

Data Capture: In order to obtain the relevant information from participants' smartphones, an android application module was developed using the **Eclipse Integrated Development Environment** (IDE). This was later integrated into AMIZONER, as explained earlier, which is hosted on the **Google Play Store**.

Transfer Protocol: A server was required to store and aggregate the data collected. **Parse.com**, a Mobile Backend as a Service (MBaaS), was chosen.

Data Aggregation: The information collected on Parse.com was exported and converted to CSV (Comma Separated Values) files using **Java Programs** developed specifically for the post-collection processing of data. Many fields were reformatted into a more acceptable form. For Example, Age was computed from one's date of birth and Telecom Service Provider Names like Airtel was computed from the Mobile Network Codes (MNC) numbers. The aggregated information has been further organized using **Microsoft Excel**.

For applying data mining techniques like Apriori Algorithm, **WEKA** (Waikato Environment for Knowledge Analysis) Software has been used.

Data Visualization: Graphs and Charts have been generated using online graphing service like **Google harts** and **Plotly**.

E. Study Limitations

Although there are undeniable advantages to using an app to collect the data, it has certain limitations. The Calling, SMS and Mobile internet information tend to be incomplete or inconsistent during the following cases:

- Students use devices without SIM (like Tablets), Dual SIM Smart Phones, Blackberry 10 OS Phones or Emulators (like BlueStack allowing one to use android Apps on a Desktop)
- Student has a habit of deleting the messages in his inbox or sent folder, the SMS information collected may misrepresent the user observed.
- Student has shared his mobile device with someone else. Therefore, the data collected does not represent the student's usage despite having the student's demographic information recorded.
- Network Data has been collected with variable time periods. That is, while one user's network usage may have been successfully observed for 7 days, another user's network usage may have only been observed for a few hours. In order to make it more reliable, only those

Users with at least 72 hours (3 days) of observation have been considered. To ensure more uniformity, all analysis has been made on the computed average network data used per hour per user.

These limitations have been considered and all such cases that have been successfully identified have been filtered to remove any invalid or inconsistent record.

III. DATA COLLECTION, ANALYSIS, INTERPRETATION AND RESULTS

After the filtering and removal of invalid and inconsistent data records, only 4465 records will be used for analysis. The analysis and interpretation of this data collected has been divided into 6 parts, explained in detail below. Table-1 represents the data for the same.

A. Demographic Variables

It can be seen that 95% of the sample are within the **age** of 18 to 24 years (Table-1). This clearly indicates that the sample is more representative of young college students. The minimum age recorded was of 16 years and the maximum age recorded was 52 years. Those participants over 30 years of age are currently pursuing Post Graduate Degrees like Ph.D and MBA.

Table-1 Description of Age group of sample collected

Description (Age)	Count	%Total
14 - 17	181	4.05%
18 - 24	4213	94.36%
25 - 34	69	1.55%
35 - 44	1	0.02%
45 - 54	1	0.02%

The male-female ratio (**Gender**) of the sample is roughly 3:1 where approximately 28% are females and 72% are males. **Marital Status** is also one of the demographic variables measured. An overwhelming 99.75% of the sample is single. This is understandable since the primary age group is 18-24 years. (Table-2)

Table-2 Description of Gender distribution

Description (Gender)	Count	%Total
Male	3211	71.91%
Female	1254	28.09%

Education is also collected as a demographic variable with a total of 187 unique degrees collected under 95 institutions under 6 Amity Universities across 4 adjoining States of India. About 42% of the participants are students who are currently pursuing a Bachelor of Technology Degree. (Table-3 and Table-4)

Table-3 Description of Year wise distribution

Description (Year)	Count	%Total
Year 1	1694	37.94%
Year 2	1684	37.72%
Year 3	827	18.52%
Year 4	240	5.38%
Year 5	20	0.45%

Table-4 Description of Campus wise distribution

Description (University)	Count	%Total
Amity University, Noida, (Uttar Pradesh)	2880	64.50%
Amity University, Lucknow, (Uttar Pradesh)	648	14.51%
Amity University, Gurgaon (Haryana)	580	12.99%
Amity University, Jaipur (Rajasthan)	273	6.11%
Amity University, Gwalior, (Madhya Pradesh)	78	1.75%
Amity Greater Noida, (Uttar Pradesh)	6	0.13%

B. Smartphones Devices

Popular **brands** of android smartphones as well as the most widely used models have also been analysed. SAMSUNG is the top used brand with 38% which is greater than SONY, MOTOROLA and HTC combined. 5% of the total sample used smartphones from brands that if taken individually were insignificant. These brands include INTEX, SPICE, AQUA, KARBONN, PANASONIC, etc.
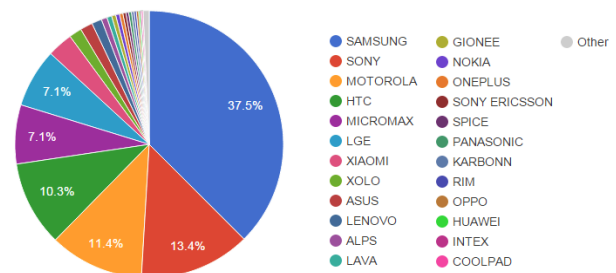

Fig.1. Smartphone Brands

Besides the mobile phone brands, the model number of each device was also collected. **Approximate prices** for each handset model have been found and graphed. It should be noted that mobile phone prices are dynamic and change every day. GSMARENA.COM, one of the top websites providing mobile handset information, simplifies this process by assigning a grade (out of 10) which corresponds to a price range. These prices are the current value for each mobile handset (and not necessarily the original purchase price by each participant). This information was parsed using a program for each handset model collected and reformatted to accommodate price ranges in Indian Rupees.

It was found that the dominant price ranges for smartphones were either within Rs 10,000 – Rs 14,000 range or within the Rs 17,000 to Rs 20,000 range. Over 60% of the sample owned smartphones below Rs 17000 while about 55% of the sample owned a smartphone priced between Rs 10,000 and Rs 20,000.

Fig.2. Price range of Mobile Handset

C. Network Service Providers

Among the **Network Operators** in India, Airtel is leading with approximately 43% of the total sample. Vodafone is close behind with approximately 38%. Together, both Airtel and Vodafone make up almost 80% of the total sample.
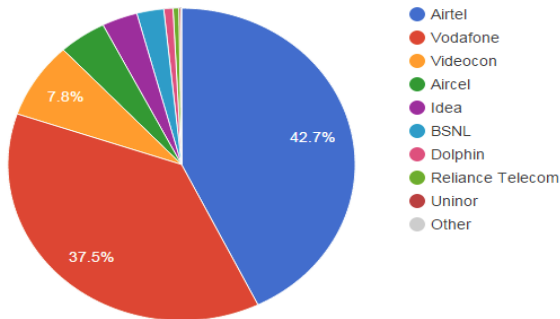

Fig.3. Network Providers

Although Airtel leads overall, Vodafone has a far greater reach in Noida and Lucknow. It was interesting to note that in Gwalior, Videocon is leading with over 40% of the total sample (whose campus location is at Gwalior) and in Lucknow, Vodafone is leading by almost 35% of the total sample (whose campus location is in Lucknow).

D. Internet Usage

We can further understand whether the user primarily uses Mobile Internet (EDGE/2G/3G/4G), or relies on free WiFi (with respect to Mobile Network Operators) or both. It can also measure which the user uses more (in terms of data volume). This has huge implications with regard to Telecom industries considering that they earn only when a user avails their internet packs. The Volume of data consumed can also be used as a basis for determining suitable packs for college students.
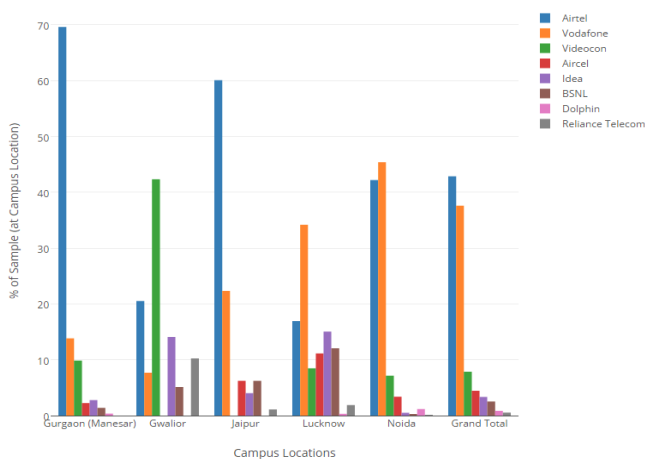

Fig.4. Mobile charges by campus location

The network data had been collected with variable time periods. That is, while one user's network usage may have been successfully observed for over 7 days, another user's network usage may have only been observed for a few hours. In order to make it more reliable, only those users with at least 72 hours (3 days) of observation have been considered. To ensure more uniformity, all analysis has been made on the computed average network data used per hour per user.

From the Area Plot in Fig.5., we can see that younger the smartphone user, greater is the reliance on WiFi. With age,

the user starts relying on Mobile Internet instead of WiFi alone. This is understood from how the gap between Mobile Data used and WiFi data used are significant at a lower age but almost the same at an older age.
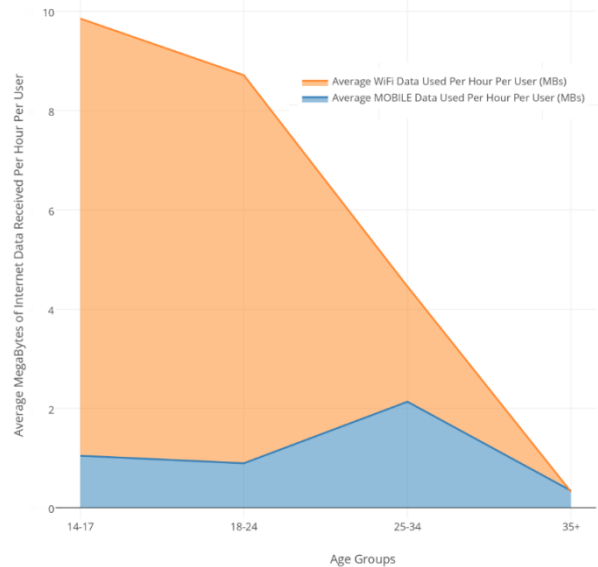

Fig.5. Area Plot: Network usage w.r.t age groups

Internet usage has also been related with two other variables: Gender and Chat App Usage. If we compare the internet usage between males and females, it is seen that males consume more Internet data (*WiFi and Mobile*) than females. More specifically, males consume 40% more Internet Data (average number of megabytes consumed per hour per day) and 29% more Mobile Internet Data than Females on their smartphones. Interesting to note is that Non-Chat App users consume more Internet Data than that of Chat App Users. Non-Chat App users use 9% more Internet data per hour per day and 36% more *Mobile* Internet Data than Chat App Users.

E. SMS and Call Usage

The first objective is to verify whether or not the number of SMS sent has dropped with the growing popularity of Chat Apps. Using the data collected, , a time series graph has been generated based on the average number of SMS sent per day per user over three years from January 2012 to December 2014. These were originally computed date-wise which have been further aggregated to represent quarters of each year. On plotting this data, it can be seen that the number of SMS sent have continued to drop over time. It's interesting to note certain irregularities in the graph. This is assumed to be because of regulations set forth around that time by Telecom Regulatory Authority of India (TRAI). On **September 2011**, TRAI limited the number of SMS that can be sent by each user to 200 per day. On **June, 2012**, the limit of 200 SMS per day was removed. On **November, 2012**, a limit of 100 SMS per day for discounted packs was introduced. However, one could send more than 100 SMS by paying higher rates for each SMS sent. All this has created a visible fall in the average number of SMS sent per day.
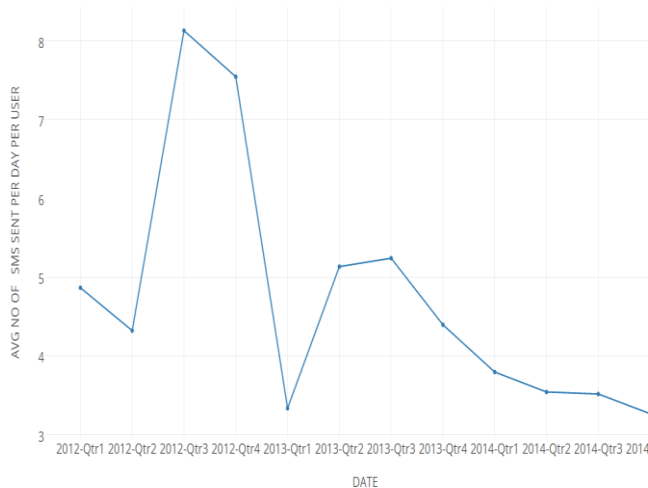
Fig.5. Time Series: No of SMS SENT (2012-2014)

On comparing Chat App users and Non-Chat App users, it was noticed that Non Chat App users send more SMS with greater number of characters than those of Chat App Users. Non-Chat App users also have greater tendency to call than Chat App Users. However, the average number of minutes spent in one call was relatively the same for both types of users.

On comparing the SMS usage between Males and Females, it was seen that Females on average send about 10 words per SMS while males send about 9 words per SMS. Females used emote in 12% of their total sent SMS while Males used emotes in only 8% of their total sent SMS. It was seen that on average, an Amity University student uses about 49 characters per SMS.

When we compare the number of SMS sent and calls made throughout the week, it is seen that both SMS and Calls vary similarly with respect to the days of the week. It was also seen that the number of Calls made are more than twice the number of SMS sent.

F. Mobile Application Usage

We can categorize the participants on the basis of whether they use Mobile Messaging Applications (Chat App Users) or not (Non Chat App Users). It was observed that an overwhelming majority (96%) of the total participants use Chat Apps.

By ranking the apps installed on the users mobile devices, WhatsApp is the most installed application (90%) with Facebook and Messenger immediately behind. In the table below, the Chat Apps among the top 20 Apps installed are bolded. Among them, WhatsApp, Facebook Messenger, Hike Messenger and ChatOn (Default SAMSUNG Chat App) are the most installed Chat Apps.

To identify the Top Apps by count is not an accurate method to evaluate the top apps used. This is because there are many chat apps that are preinstalled by the device manufacturer. For example, ChatOn was the 22nd Most Used App based on the count. However, if we were to rank the apps on the basis of the average percentage of network data sent by app, ChatOn would not even appear in the top 50 Apps (by network usage). However, if we identify the top App Categories, Communication and Social are in the top 3 of App Categories by count and by network usage. Mobile Messaging apps are usually associated with the categories Communication or Social.

It is interesting to note how categories like Shopping are in the list of Top App Categories by Network Data, but are not in the list of Top App Categories by Count. This means that although the number of Shopping Apps installed on a user's phone are less, they use up a significant amount of internet data.

The top 25 Apps ranked on the basis of % of total internet used are shown in Table 4.6.2. Facebook uses about 14% of the total Internet Data consumed by a user while Whatsapp uses about 7% of the total internet data used on average by a user. This list is unique because instead of looking at count, it ranks on the basis of internet data consumed. By selecting only the Social and Chat Apps, a bar chart has been generated identifying the top 12 Social and Communication Apps by network data used.
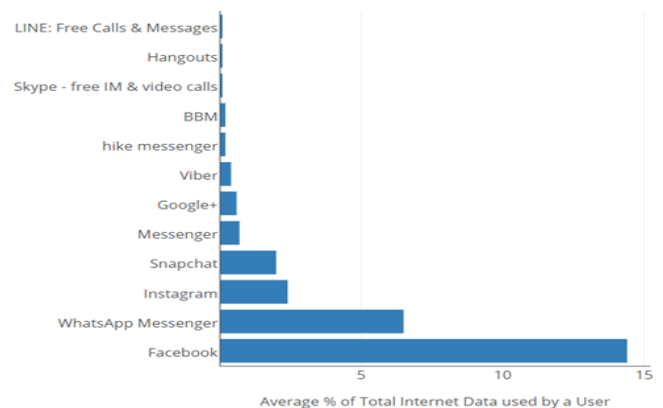

Fig.8. Top Social and Chat apps (by network data)

To further assess Chat Apps, the Apriori Algorithm has been used so identify associations among them. Using WEKA, the following conclusions have been made: 92% of those who had Facebook Messenger Installed also installed WhatsApp Messsenger while 94% of the users who had both Facebook and Facebook Messenger installed installed WhatsApp Messenger. It's clear that these 3 Apps are often installed together. Interestingly, 93% of those who had Facebook installed also installed WhatsApp Messenger while only 88% of those users, installed Facebook Messenger. WhatsApp Messenger is clearly more popular than Facebook Messenger, even among Facebook Users. Among WhatsApp Messenger Users, only 73% also installed Facebook Messenger and only 66% installed Facebook.

One interesting observation was how closely TrueCaller was associated with Facebook and WhatsApp. It was noticed that among the users of TrueCaller, 94% also used WhatsApp Messenger, 78% used Facebook Messenger and 73% used Facebook.

IV. SUMMARY, FINDINGS, SUGGESTIONS AND RECOMMENDATIONS

By using an application to track mobile usage, the study is significantly more accurate than previous studies utilizing questionnaires. Instead of measuring one's perception of high and low, the study was able to quantify each user's SMS, CALL, INTERNET and APPLICATION usage.

By doing so, it was possible to identify and assess a number of factors. This is with respect to over four thousand Amity University Students currently residing in India.

It's clear from reports (like by Portio Research and InMobi) that SMS is being replaced by Chat Apps. Chat Apps like Skype and Viber are supporting the feature to call over the internet. These VoIP (Voice over Internet Protocol) applications pose another threat to Telecommunication Providers, as Chat Apps would eventually replace even calling. Recently, BhartiAirtel Ltd planned to charge clients extra for Internet communication services such as Skype as a strategic move to both discourage and monetize off calling by Chat Apps. However, they dropped the plan.

From the analysis however, it was seen that the frequency of calls by both Chat App and Non-Chat App users were relatively the same. Also, while 90% of the sample had Whatsapp installed, only 18% had Viber installed and 1.5% had Skype installed. As Whatsapp is the most used mobile application for text messaging, Viber is one of the most used VoIP mobile applications. This shows that Calling via Chat Apps is not as frequently used as messaging via Chat Apps. That is, Calling via service providers is still preferred over Chat Apps by the majority.

On the basis of number of applications installed, the top 3 Mobile App Categories were Social (30% of all running apps), Communication (20% of all running apps) and Tools (11% of all running apps). Together, Chat Apps fall under one of the two categories, Social or Communication. This indicates the categories App Developers should focus on. Also, since over 71% of the sample use Facebook, App Developers should consider Facebook Advertising as a reliable means to promote their applications.

REFERENCES

1. Pempek, T.A, Yermolayeva, Y.A and Calvert, S.L (2009) "College students' social networking experiences on Facebook", Journal of Applied Developmental Psychology
2. InMobi (2014,August). Mobile Messaging App Study
3. Walton, N. (2014). Chat apps will disrupt global telecoms sector. The Oxford Analytica Daily Brief.
4. Gomathi, C., & Gowtham, P. V. (2013). Social Media Networking State of Social Media. In Proceedings of National Conference on New Horizons in IT-NCNHIT (p. 205).
5. Church, K., & de Oliveira, R. (2013, August). What's up with whatsapp?: comparing mobile instant messaging behaviors with traditional SMS. InProceedings of the 15th international conference on Human-computer interaction with mobile devices and services (pp. 352-361). ACM.
6. Portio Research (2014, March). Mobile Messaging Markets 2014: Facebook, Whatsapp, SMS and OTT
7. El-Darwiche, B., Friedrich, R., Péladeau, P., & Sabbagh, K. (2010). The Future of Telecom Operators-Capabilities for Rapid Change.
8. Do, T.M.T and Gatica-Perez D. (2010). By their apps you shall understand them: mining large-scale patterns of mobile phone usage. In Proc.MUM, pages 27:1–27:10, Limassol, Cyprus, 2010.
9. Battestini, A., Setlur, V., & Sohn, T. (2010, September). A large scale study of text-messaging use. In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services (pp. 229-238). ACM.
10. Falaki, H, Mahajan R, Kandula S, Lymberopoulos D, Govindan R, and Estrin. D (2010). Diversity in smartphone usage. In Proc. MobiSys, pages 179–194, San Francisco, California, USA, 2010.
11. Bernhard, K., Pretel, I., Vanhecke, S., Vilariño, A. B. L., & de Ipiña, D. L. (2013). Analysis of log file data to understand mobile service context and usage patterns. IJIMAI, 2(3), 15-22.
12. "More Than a Quarter of the World Will Use Mobile Messaging Apps by 2019" , eMarketer's, November 29, 2016. Available at: https://www.emarketer.com/Article/More-Than-Quarter-of-World-Will-Use-Mobile-Messaging-Apps-by-2019/1014773

AUTHORS PROFILE

Bhopendra Singh is working as a Head of Industry Relations-Engineering and Senior Faculty in the Department of Engineering and Architecture, Amity University Dubai. His research interests include optical communication, Antireflection coating for solar cell, Synthetic grating SOT structure, Sports Science, RFID control system. He has attended workshops, seminars and conferences related to Artificial intelligence and he has also published many papers in reputed journals, authored books and filed three patents.

Vinod Kumar Shukla is working as a capacity of Faculty for Information Technology in the Department of Engineering and Architecture, Amity University. His research interests include Internet of Things, Information technology, Semantic Web, RFID control system. He has attended many workshops, seminars and conferences related to Information technology and he has also published various papers on the listed areas.

Sonia Singh fully blended professional with vast experience in Industry, training, teaching and consulting in India MENA region. Presently associated with many companies as consultant and help them to achieve business excellence. Also, she is guest faculty with Al Falah University, Dubai in the area of finance and banking. She is founder of Toss Global Management, an organization for education and management consultancy. She is also honorary advisor of Gulf American University.