Hindawi
Security and Communication Networks
Volume 2021, Article ID 7940103, 21 pages
https://doi.org/10.1155/2021/7940103

Research Article

S²NOW: Secure Social Network Ontology Using WhatsApp

Rahul Johari,[1] **Sawan Kalra**,[1] **Sonika Dahiya**,[2] **and Kalpana Gupta**[3]

[1]*SWINGER: Security, Wireless, IoT Network Group of Engineering and Research Lab, University School of Information, Communication Technology (USICT), Guru Gobind Singh Indraprastha University, Sector-16C, Dwarka, Delhi, India*
[2]*Department of Computer Science and Engineering, Delhi Technological University, Delhi, India*
[3]*Academics Division, Centre for Development of Advanced Computing (C-DAC), Noida, India*

Correspondence should be addressed to Rahul Johari; rahuljohari@hotmail.com

Received 21 January 2019; Accepted 28 December 2020; Published 31 January 2021

Academic Editor: Rohit Ranchal

Communicating and sharing the ideas and feelings between human beings form the basis of social network. With the advent of different social network tools and applications, the task of networking between people has reached to next level, crossing bridges and boundaries. However, is there a connection or linkage between the frequent and sometime just random exchange of text messages between people using social network-based tools and applications and also are these messages secure? Or are they vulnerable to threat and attacks? Literature survey shows that very little work has been done in this direction. With this intention, in the current research work, a real-time case study has been taken identifying the creation of the social network between a group of persons who frequently chat using the WhatsApp Messenger application based on their common hobbies, choices, and areas of interest and applying cryptographic techniques to ensure the security of data (chats). The proposed study uses the Caesar Cipher Cryptographic Technique and newly proposed Block Quadra Crypto Technique for encrypting the chats and showcases a comparison between the two techniques. The results are encouraging and exceed the expectation.

1. Introduction

Social network is the network that get established when a group of people share common interest with each other, bond with each other, share the feelings and sentiments with each other, and connect with each other digitally towards sharing common and mutually acceptable thoughts, feelings, and emotions. This is a reason that why various social networking platforms such as Facebook, WhatsApp, Instagram, Vine, Tumblr, Meetup, and ClassMates all have bombed our social life. In today's digital world, bonding on social platform is quite common. Every day, every person spends quality time of his life by devoting hour(s) on these social networking sites, tools, and applications (STA) searching for old peers, new friends, and acquaintances. The bonding happens naturally between the people that have some common point-of-interest (PoI) such as connection between individuals that are a part of college football clubs, movie and theatre societies, career-oriented women (CoW), and recreation clubs. With more digitization going to happen in the coming years, these sites, tools, and applications are here to stay and would definitely impact the way we live, think, express, and bond with fellow human beings. In this current research work, an effort has been made to identify the social network link by taking into consideration the contents of the messages and chat exchanged on the WhatsApp Messenger application of different persons for the duration of one month (or 30 days). The study not only tries to detect connections between chats and formulate a social inclination between the selected sample size but also provide security to the chats by applying encryption techniques to the chats, thereby ensuring data privacy. In the current research work, the two encryption techniques that have been used for securing chats are Caesar Cipher and the newly proposed Block Quadra Crypto Technique. A comparison is also showcased between the two encryption techniques.

Cryptographic encryption is of two types: symmetric encryption and asymmetric encryption. Symmetric encryption in turn can be implemented through cryptographic

techniques, which include techniques such as Caesar Cipher, Vigenere Cipher, Monoalphabetic Cipher, Polyalphabetic Cipher, and Transposition Cipher, and through cryptographic algorithms, which include RSA, DES, and AES algorithms.

To show the effectiveness of the proposed work, the crypto technique has been coded in C# Programming Language in the front end whereas MongoDB has been used in the back end as it is a vital component of MEAN Software Stack.

The novelty of the work is that, to the best of our knowledge, understanding, and wisdom, the current study is first of its kind where a social networking tool such as WhatsApp has been undertaken as a single source of study where end-to-end user chats are made safe and secure using symmetric cipher techniques and newly postulated Block Quadra Crypto Technique. Not only this, the study focuses on the integration and mapping of WhatsApp with the Big Data tools and technologies, which were always viewed separately and never have been integrated in the past. It is believed that the security aspect of proposed work would throw challenges to the research community leading further pathway to the devise and develop the new, effective, and efficient cryptographic techniques.

For the sake of clarity and simplification, the rest of the paper is organised as follows: Section 2 describes the related work on social media data mining. Section 3 describes the motivation for pursuing current research work, Section 4 describes the problem statement, Section 5 illustrates the methodology adopted, Section 6 describes the procedure that was adopted, followed by the flowchart and mathematical modeling of the S^2NOW algorithm, Section 7 describes the results obtained, Section 8 describes the conclusion and future work, and Section 9 describes acknowledgment followed by references.

2. Related Work

In [1], authors have presented elaborated details about the advantages and disadvantages of Big Data tools and techniques. In [2], authors have presented a workflow to bring together collectively both qualitative analysis and important data mining techniques. In [3], authors commented that it is challenging for traditional databases following hierarchical and network model approach to modify and then organize the data in a form that the web user generates on daily basis while performing social activities such as online-chatting, liking, poking and pinging, and tweeting on social media which over period of time have become a huge source of semistructured and unstructured data. In [4], authors have illustrated a framework to improve the throughput of the system. In this paper, two issues related to quick access of real-time data are presented; first issue deals with indexing, and the second one deals with handling of data. Based on these concepts, RTSEs (real-time search engines) were analysed.

In [5], authors comment that students of Polytechnics in Ghana are dependent on the WhatsApp application for their daily exchange of messages as compared to using mobile voice calls or phone calls. Economical, fast, easy, and effective modes of communication are some of the prominent features that influenced the decision of the students. To illustrate it further, a student survey was carried out from January 2014 to June 2014 where a questionnaire was designed to compile data from a sample of 600 odd students from three polytechnic institutions of Ghana. The study revealed the students' attention in switching to WhatsApp application as their most preferred mode of social network tools and applications.

In [6], authors have analysed the performance of numerous algorithms such FCM, IFCM, and T2FCM on noise free and noisy data (also linearly and on nonlinearly separable data). In [7], the author describes how the usage of WhatsApp in smart phones has led to significant effect on writing, usage of vocabulary word choice, and voice of Saudi students studying in College. In [8], authors describe a study that seeks to empirically identify the impact and magnitude of social network application such as WhatsApp Messenger on the overall performance of the students in Ghana. In [9], authors focus on the use of mobile phones as a platform to pursue high-quality collaborative and coordinated research work amongst a selected group of third-year UG Computer Science students. The students used a WhatsApp application to have collaborative group conversations about their ongoing research project. In [10], the author describes the effectiveness of using mobile tools and technologies and STA such as WhatsApp to show how mobile learning activities are guided by activity theory using students' knowledge management system.

In [11], authors discuss the usage of WhatsApp in the effective learning of a group of 37 students of B. Ed. (Bachelor in Education) programme during session 2013-2014. In [12], authors conducted an extensive study to explore and understand classroom interaction between teachers and students of school using WhatsApp. To achieve it, the authors carried out twelve half-structured interviews with faculty members who use the WhatsApp application in order to effectively interact with their students. In [13], authors compared the communication practices done by the users using textese and other ways of online writing in WhatsApp across two eras of generations. The authors comment that failing to communicate using a variety of languages cannot be systematically associated with their communicative practices. In [14], the authors examined the undergraduate students at Universiti Brunei Darussalam depicting the convenience of WhatsApp in their everyday lives and some critical issues arising with it. In [15], authors described how the texters mobilize transcribed laughter (i.e., hahaha) by focusing on the position of smiled and laughter in the message and its course of actions including unilateral laughter and when the user laughs before producing a message. In [16], the authors showcased how information was spread through WhatsApp during the search and rescue operation of three youths kidnapped and detected that 9 out of the 13 rumours circulating on WhatsApp were true.

3. Motivation

WhatsApp is a social network based online instant messaging application that runs on a variety of heterogeneous operating systems such as BlackBerry, Android, or iOS. As a

result, the motivation of undertaking the current research work is to mine the conversations and figure out social interests between individuals as a major time of everyone's life is now spent on communicating these messages on STA. The social conduct of the individuals coupled with the topics of interest and the intensity/gravity or the seriousness of the topics provided valuable information to establish an understanding of the customer base who extensively uses STA such as WhatsApp. This could also find application in the corporate world as official chats can be analysed for social interests or sentiment that brings together certain category of corporate honchos together in social groups. This unexplored area acted as motivation to work in this direction and attain the plausible results. The current exchange of chats over STA is majorly insecure, often vulnerable to threats and attacks. Recent work performed by security experts highlights the case as in [17] as per the market survey mobile-based communication apps and platforms such as Skype and WhatsApp have far less or almost no protection against app-based phishing as compared to e-mail.

In [18], researchers comment that all the users who access web versions of WhatsApp are likely to get cyber-attacked. They have detected a vulnerability which has not been patched yet and that could allow hackers to gain control over end-user accounts, thereby gaining access to personal data using a malware-laced image resulting in customer privacy violation.

As per [19], the WhatsApp suffered a massive cyber-attack resulting in downtime of around 120 minutes affecting more than 175 million users. As per [20], software architects pointed out various bugs such as "quote" feature in a group-oriented conversation to change the identity of the source, even if that individual may not be a member of that particular social group or modifying the text of someone else's reply or transmitting a private message to fellow cogroup participant, being disguised as a full-view public message for everyone, so when the targeted individual replies back, it is visible to all in the conversation.

In [21], new research from Check Point reveals that WhatsApp vulnerabilities enable attackers to manipulate, tamper, and modify messages (MTM) both in public and private conversations. This type of MTM usually makes it easy for intruder to spread fake messages and misinformation. In [22], authors claim that engineers and researchers detected three prominent WhatsApp vulnerabilities that enable users to not only intercept and intrude but also manipulate messages in private and group chats. These WhatsApp vulnerabilities enable hackers to tamper messages, launch vicious online scams, and spread artificial rumours and fake news resulting in what is now popularly called as "FakesApp."

In [23], technical writers' comments that Facebook owned WhatsApp has major problem with the widespread circulation of fake news and rumours. With a result, they had to put control and restrictions on the number of times a specific message could be forwarded to the other users on the group.

This acted as a great zeal of inspiration to work in this direction, and the newly proposed Block Quadra Crypto Technique is the outcome of the same.

4. Problem Statement

In the current social networking scenario, no research work has been done till date on social networking application "WhatsApp," leading to its linkage or connection to social network as no authentic reports or publications could be traced on the web. The current research work focuses on social networking in WhatsApp, providing security to the chats using symmetric cipher technique that means the study focuses on the integration and mapping of WhatsApp with the Big Data tools and technologies, which were always viewed separately and had never have been integrated in the past.

Will the social networking through WhatsApp provide intriguing results upon which various graphical analytics can be applied and whether the graph having nodes/vertices representing a person's WhatsApp chat and edges that define how strong a connection between two persons or nodes can be plotted? Also, in the graph, the weights on the edges increase as the number of common words exchanged between two chats (nodes) increases, and does this hypothesis hold true in all the environments? This is something that needs to be ascertained.

5. Methodology Adopted

5.1. Mapping Table. Proposed Table 1 is used for conversions that were required in securing the chats both while encrypting as well as decrypting.

5.2. Caesar Cipher. In the field of cryptography, it is one the naive and simplest encryption technique. It is an additive cryptographic technique. Each letter of the plain text is replaced. The cryptographic equation followed, here, is

$$\mathbf{C} = (\mathbf{P} + \mathbf{K}) \bmod 95, \tag{1}$$

where $\boldsymbol{C}$ is the cipher text, $\boldsymbol{P}$ is the plain text, $\boldsymbol{K}$ is the key (**3**), and "95" is the total number of characters in the mapping table.

For example, the plain text is "This is in Caesar Cipher."

After the application of Caesar Cipher, the cipher text obtained will be as follows: "Wklv#lv#lq#Fdhvdu#Flskhu" (an addition of 3 letters in the plain text and the new positional value according to the mapping table, as listed in Table 1).

For instance, character "T" has the positional value 52 in the mapping table.

So, $(52 + 3) \bmod 95 = 55 \bmod 95$.

$55 \bmod 95 = 55$.

Value at 55 in the mapping table is "W." So, "T" will be replaced by "W."

Similarly, character "}" has the positional value 93 in the mapping table.

So, $(93 + 3) \bmod 95 = 96 \bmod 95$.

$96 \bmod 95 = 1$.

Value at 1 in the mapping table is "!." So, "}" will be replaced by "!."

Table 1: Mapping table.

Code	Value	Code	Value	Code	Value
0	SPACE	37	E	74	J
1	!	38	F	75	K
2	“	39	G	76	L
3	#	40	H	77	M
4	$	41	I	78	N
5	%	42	J	79	O
6	&	43	K	80	P
7	’	44	L	81	Q
8	(	45	M	82	R
9	)	46	N	83	S
10	*	47	O	84	T
11	+	48	P	85	U
12	,	49	Q	86	V
13	-	50	R	87	W
14	.	51	S	88	X
15	/	52	T	89	Y
16	0	53	U	90	Z
17	1	54	V	91	{
18	2	55	W	92	|
19	3	56	X	93	}
20	4	57	Y	94	~
21	5	58	Z	—	—
22	6	59	[	—	—
23	7	60	\	—	—
24	8	61	]	—	—
25	9	62	^	—	—
26	:	63	_	—	—
27	;	64	`	—	—
28	<	65	A	—	—
29	=	66	B	—	—
30	>	67	C	—	—
31	?	68	D	—	—
32	@	69	E	—	—
33	A	70	F	—	—
34	B	71	G	—	—
35	C	72	H	—	—
36	D	73	I	—	—

5.3. Block Quadra Crypto Technique. In this newly proposed technique, additive shift cipher is applied to the plain text and then these are read character by character, and each character’s binary equivalent value is fed into different quadrants sequentially.

For example, the plain text is “This is Block Quadra Crypto Technique.”

After the application of additive shift, the text will be as follows:

“rs}*s}∗Lvymu*[kn|k*M|$z ~ y*omrxs{ o.”

Now, the text obtained is fed into 4 different quadrants as shown in Figure 1 (in binary) and Figure 2 (in characters).

Since the distribution provides an unequal number of characters in each quadrant, a space is padded at the end of each quadrant (Table 2).

The values in these quadrants will be as follows:

1st quadrant:

00111110,00001010,00101100,01010101,01001011,00001010,01011010,00111110,01011000 and 01001111 (which are equivalent to ^, *, L, u, k, *, z, ^, x and o).

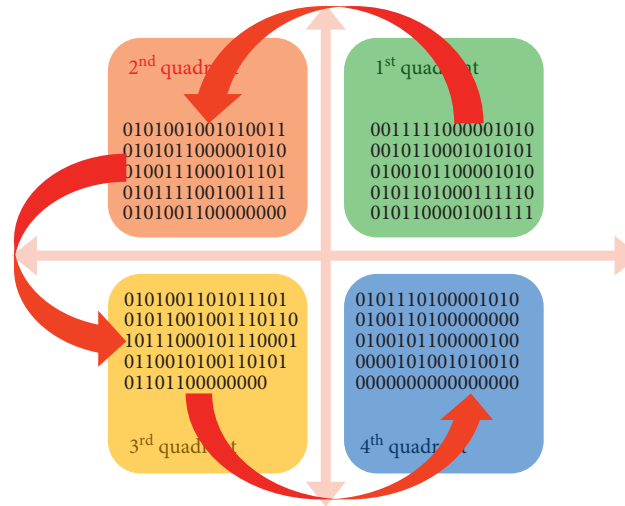


Figure 1: Depiction of the values (in binary) in each quadrant.

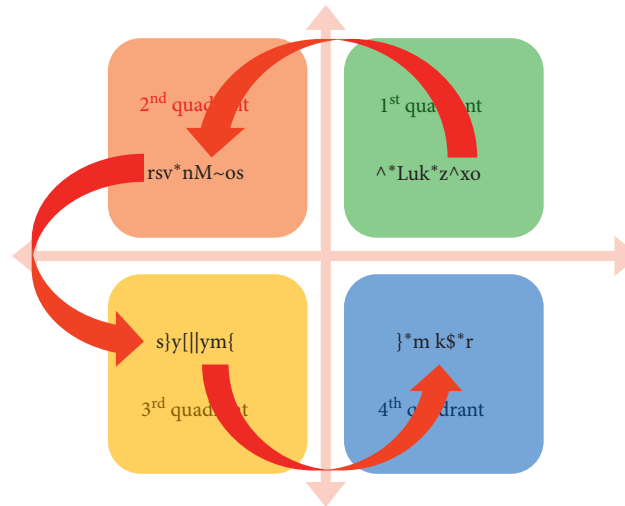


Figure 2: Depiction of the values (character equivalent) in each quadrant.

2nd quadrant:

01010010, 01010011, 01010110, 00001010, 01001110, 00101101, 01011110, 01001111, 01010011 and 00000000. (which are equivalent to r, s, v, *, n, M, ~, o, s and “).

3rd quadrant:

01010011, 01011101, 01011001, 00111011, 01011100, 01011100, 01011001, 01001101, 01011011 and 00000000. (which are equivalent to s, }, y, [, |, |, y, m, { and “).

4th quadrant:

01011101, 00001010, 01001101, 00000000, 01001011, 00000100, 00001010, 01010010, 00000000 and 00000000 (which are equivalent to }, *, m,“, k, $, *, r,“and “).

Now, the cipher text will be the contents of each quadrant read sequentially from 1st quadrant to 4th quadrant, that is, “00111110000010100010110001010101010010110000101001011010001111100101100001001111010100100101001101010110 0

Table 2: An example depicting the distribution of characters into 4 different quadrants.

Q1	^	*	L	u	k	*	z	^	X	o
Q2	R	S	V	*	n	M	~	o	S	SPC
Q3	S	}	Y	[	|	|	y	m	{	SPC
Q4	}	*	M	SPC	k	$	*	r	SPC	SPC

^	*	L	U	K
00111110	00001010	00101100	01010101	01001011
*	Z	^	X	O
00001010	01011010	00111110	01011000	01001111

r	s	V	*	N
01010010	01010011	01010110	00001010	01001110
M	~	O	S	SPC
00101101	01011110	01001111	01010011	00000000

s	}	y	[	|
01010011	01011101	01011001	00111011	01011100
|	y	m	{	SPC
01011100	01011001	01001101	01011011	00000000

}	*	m	SPC	K
01011101	00001010	01001101	00000000	01001011
$	*	r	SPC	SPC
00000100	00001010	01010010	00000000	00000000

000101001001110001011010101111001001111010100110000000010100110101110101011001001110110101110001011100101100101001101010110110000000001011101000010100100110100000000010010110000010000001010010100100000000000000000" (which is equivalent to "*Luk*ẑxorsv*nM ~ os s}y[||ym { }*m k$*r, " when mapped from the mapping table).

5.4. Mechanism Adopted. WhatsApp chats of different individuals for a duration of one month (March, 2018) were taken as part of study. The chats were obtained in text file format, that is, for each WhatsApp chat, a different text file was obtained. These chats were directly imported into Microsoft Excel application, wherein each and every chat was splitted into different columns such as time, date, and message. Then, the file was saved in .CSV (comma-separated values delimited) file format.

Now, after saving the WhatsApp chats in CSV format, each file was encrypted using a C# program, which encrypted the text in the WhatsApp messages. The chats were encrypted through both the encryption techniques and in an entirely new file.

After the application of the encryption, all the chats including original and encrypted were imported into the database, namely, MongoDB using the command (as depicted in Figures 3–5): mongoimport -d database -c collection --type .CSV --file filename.csv –-headerline.

After these chats were imported in the MongoDB database, these chats were fetched through a program, coded in C# Programming Language.

Now, these encrypted messages were received, decryption was applied on them, and the chats were obtained back in its original text or plain text form.

After the chats were obtained, analysis was performed over these and a graph was made which depicted the social inclination between all these people and along with this time taken by both the encryption techniques was compared.

6. Procedure

The procedure adopted for pursuing current research work has been showcased with the help of Figures 6–15, which are detailed as follows:

(1) Open WhatsApp and Select Options > More > Email chat (Figure 6).

(2) The text file received as attachment in the mail, Figure 7.

(3) Open Microsoft Excel and select the Data tab and click on From Text. Select Delimited as shown in Figure 8.

(4) Follow the steps in the wizard and select Comma and Other (specify "-") (Figure 9 followed by Figure 10).

(5) Now, the data looked as shown in Figure 11.

(6) Final sheet obtained was as shown in Figure 12.

(7) Save this spreadsheet in .CSV (Comma delimited) file format (Figure 13).

(8) The steps as detailed with the help of screenshots from Figures 6–13 were repeated for every chat, and a separate .CSV file was saved for each chat.

(9) These chats were encrypted in Caesar Cipher and Block Quadra Crypto Technique and imported into the MongoDB database. The chats when fetched from the database were in the form as shown in Figures 14 and 15 (Block Quadra Crypto Technique).

6.1. Algorithm Formulated. Text analytics (text mining) indicate the extraction of the information from textual data. Information extraction (IE) techniques extract/fetch structured data from unstructured text. To achieve it, the S^2NOW algorithm has been devised, S^2NOW algorithm, Secure Social Network Ontology using WhatsApp; in this, a framework/strategy is presented and results are showcased that show how a popular social media tool such as WhatsApp chat can relate two or more people by their common interest on the basis of words they had used in their respective chat. These are the various steps that are involved in the process of data mining from data collection to decision making.

Sequential procedure adopted to understand the working of the S^2NOW algorithm is detailed as follows:

Figure 3: Snapshot showing the importing of original WhatsApp chat file (.CSV) in MongoDB.

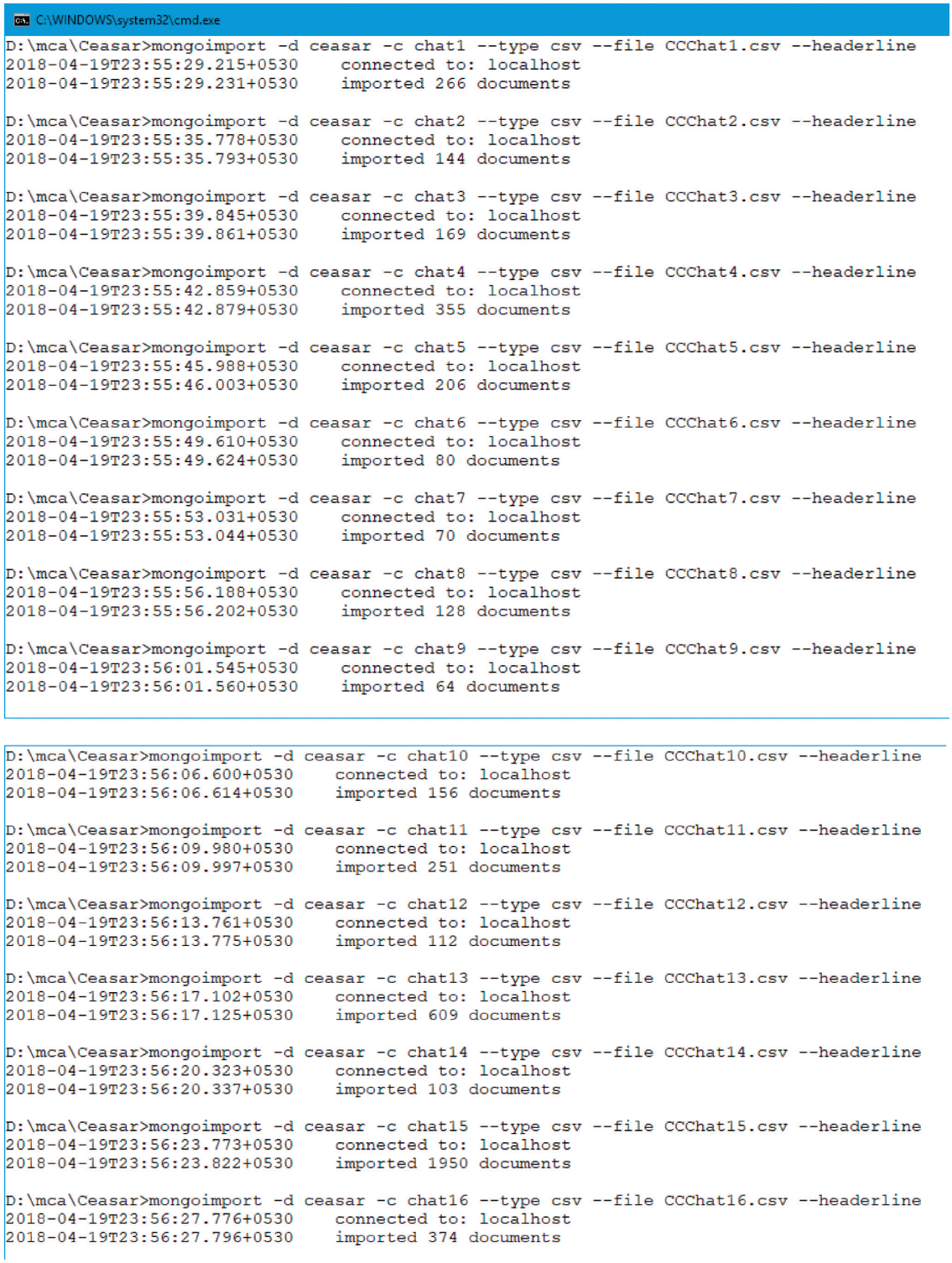


Figure 4: Snapshot showing the importing of Caesar Cipher encrypted chat file (.CSV) in MongoDB.

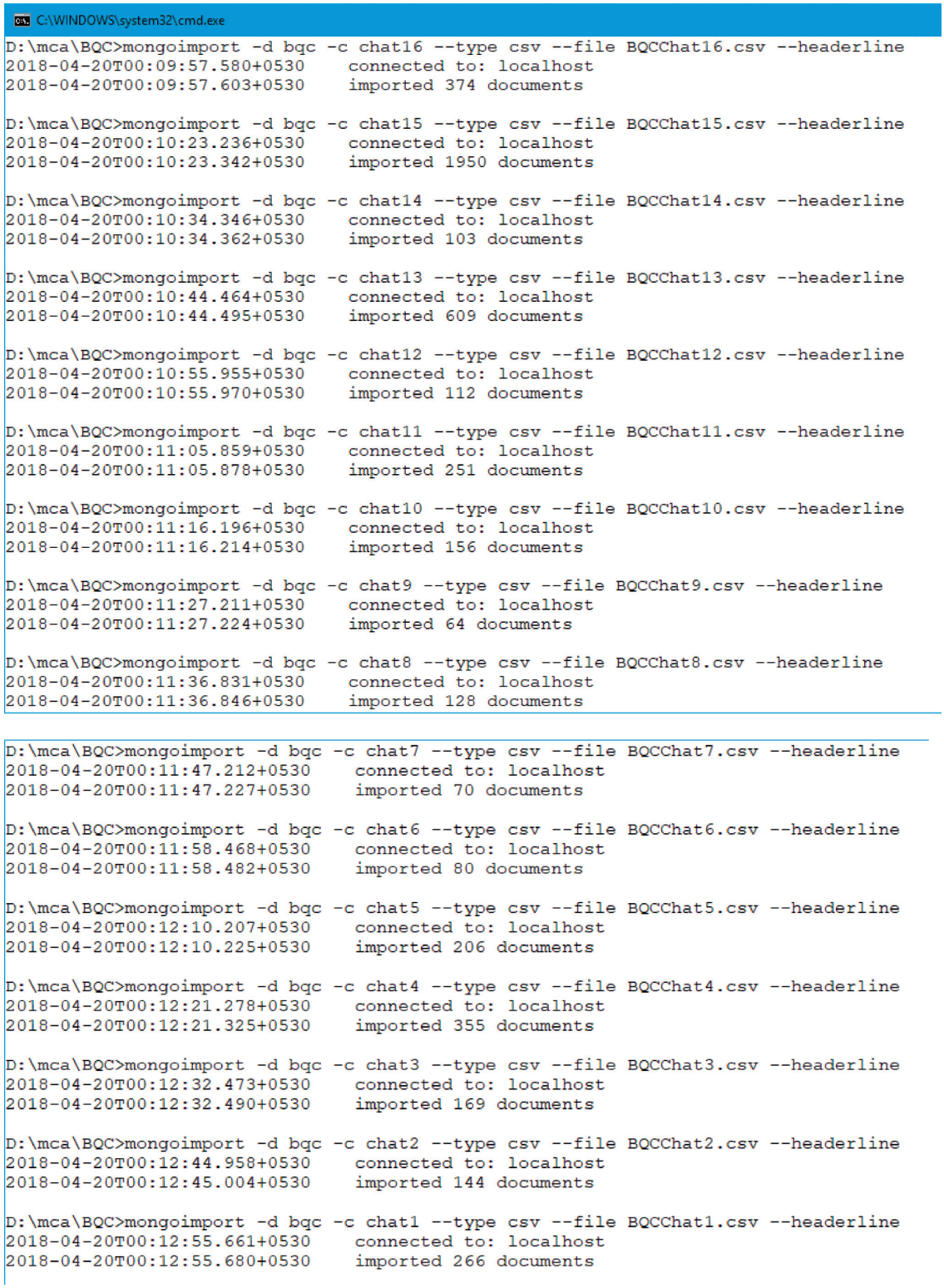


Figure 5: Snapshot showing the importing of Block Quadra Crypto encrypted chat file (.CSV) in MongoDB.

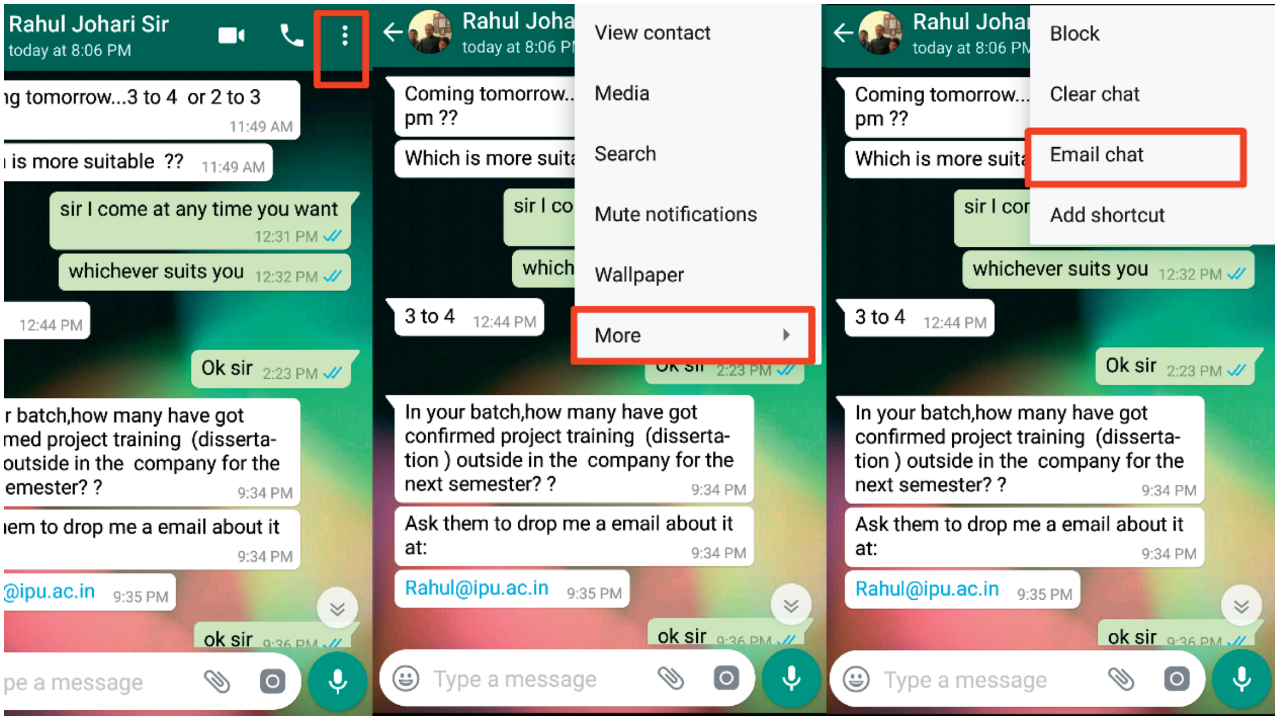


Figure 6: Steps to be followed on WhatsApp to get the data.

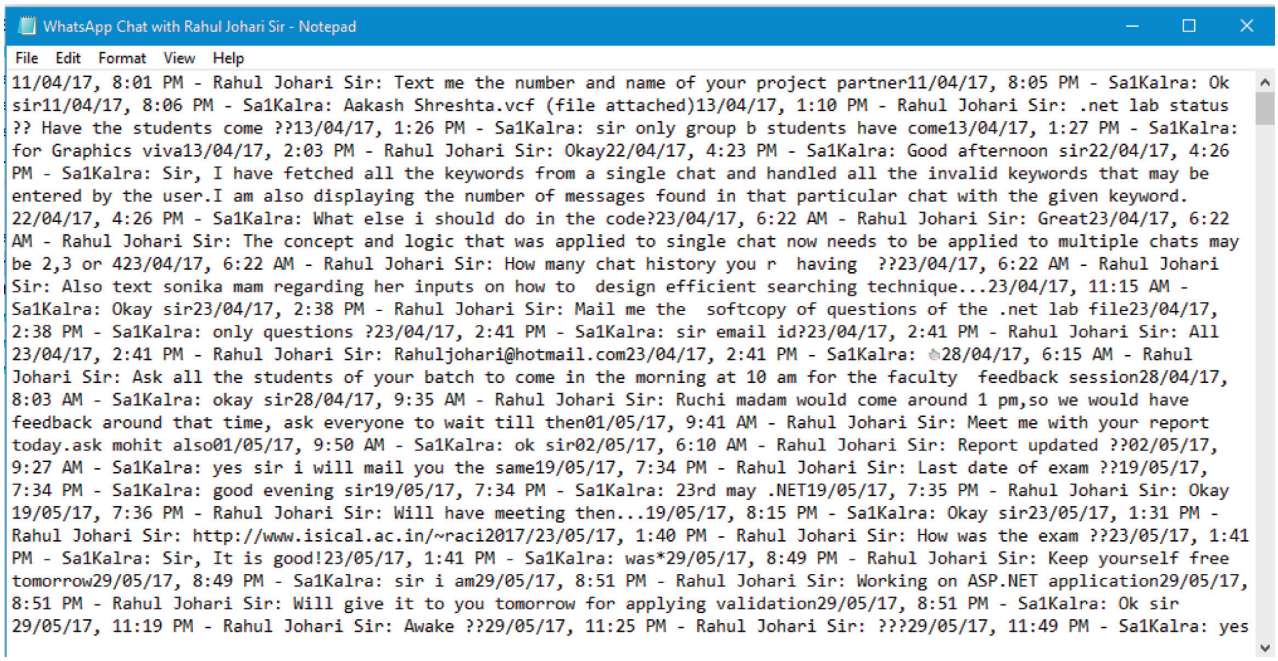


Figure 7: Contents of the text file obtained.

(1) Select the WhatsApp chat of 16 persons for a defined date and time, and let us say of 1-month duration

(2) Encrypt the chats using Caesar Cipher and Block Quadra Crypto Technique

(3) Start the MongoDB database server

(4) Fetch and decrypt the chats

(5) Take a WhatsApp chat

(5.1) Iterate/start scanning from first word to the EOF

(5.2) To remove frequently used English language connecting words such as I, we, and, if, are, and if

(5.3) To ignore all the words whose word length $(W_L) <= 6$

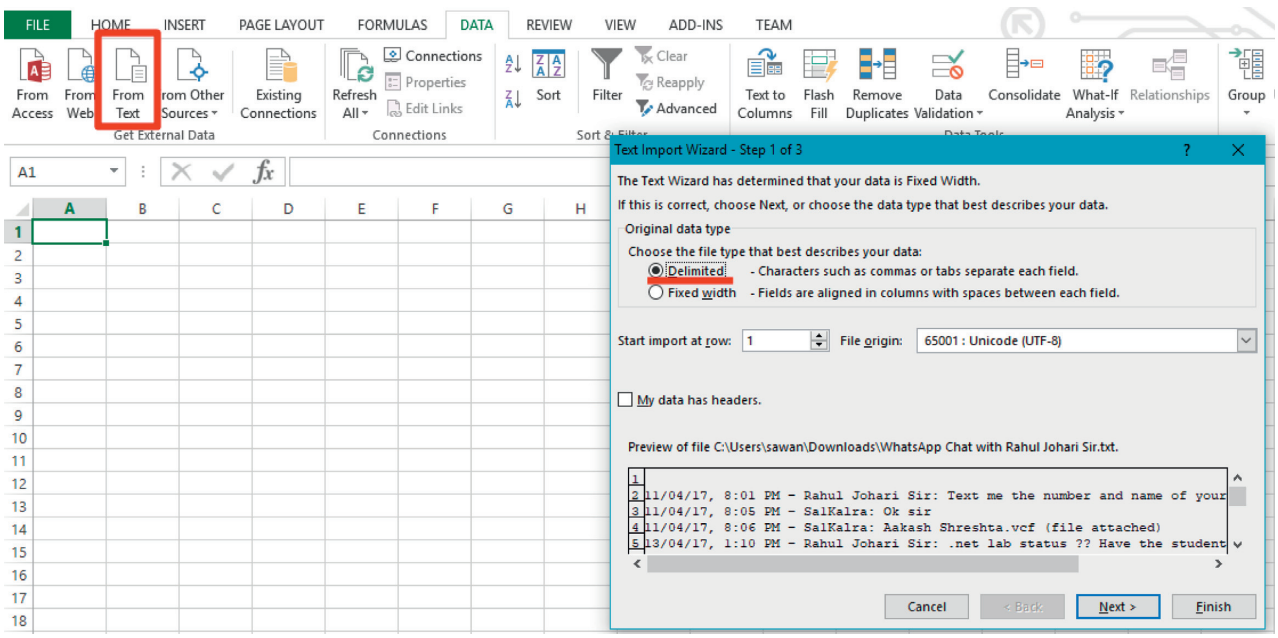


Figure 8: Importing of the text file into MS Excel.

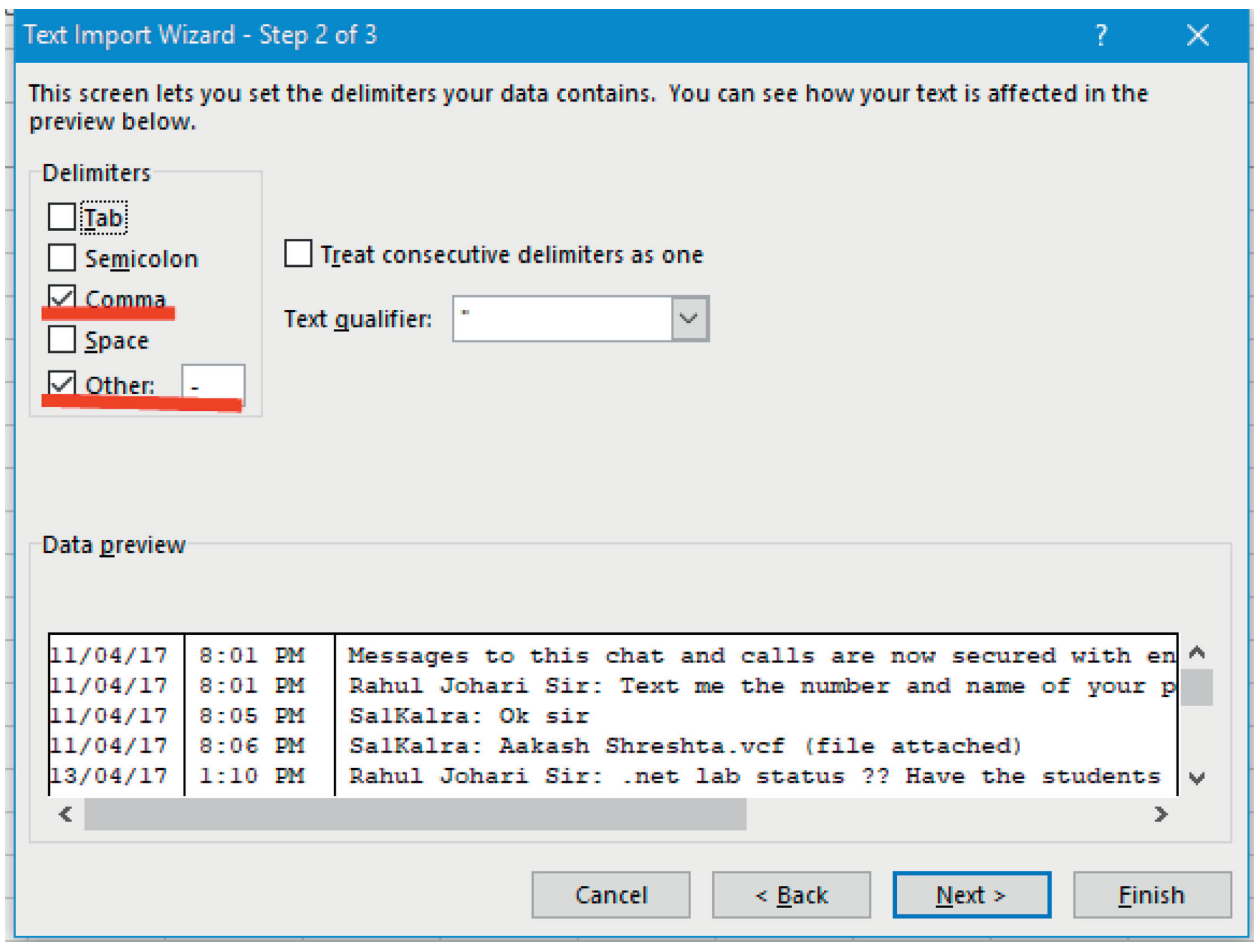


Figure 9: Text Import Wizard, selecting Comma and "-."

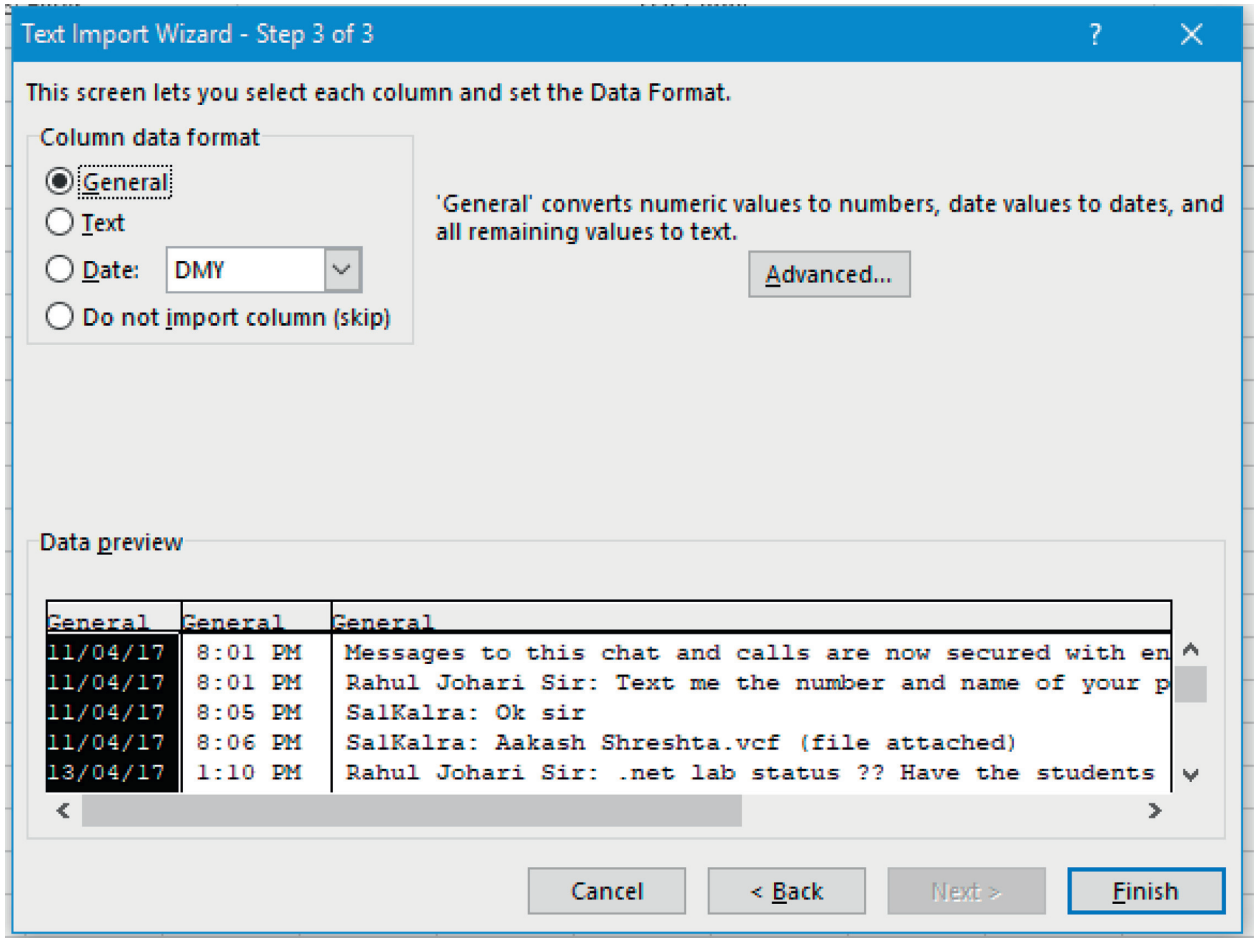


Figure 10: Text Import Wizard (step 3 of 3).

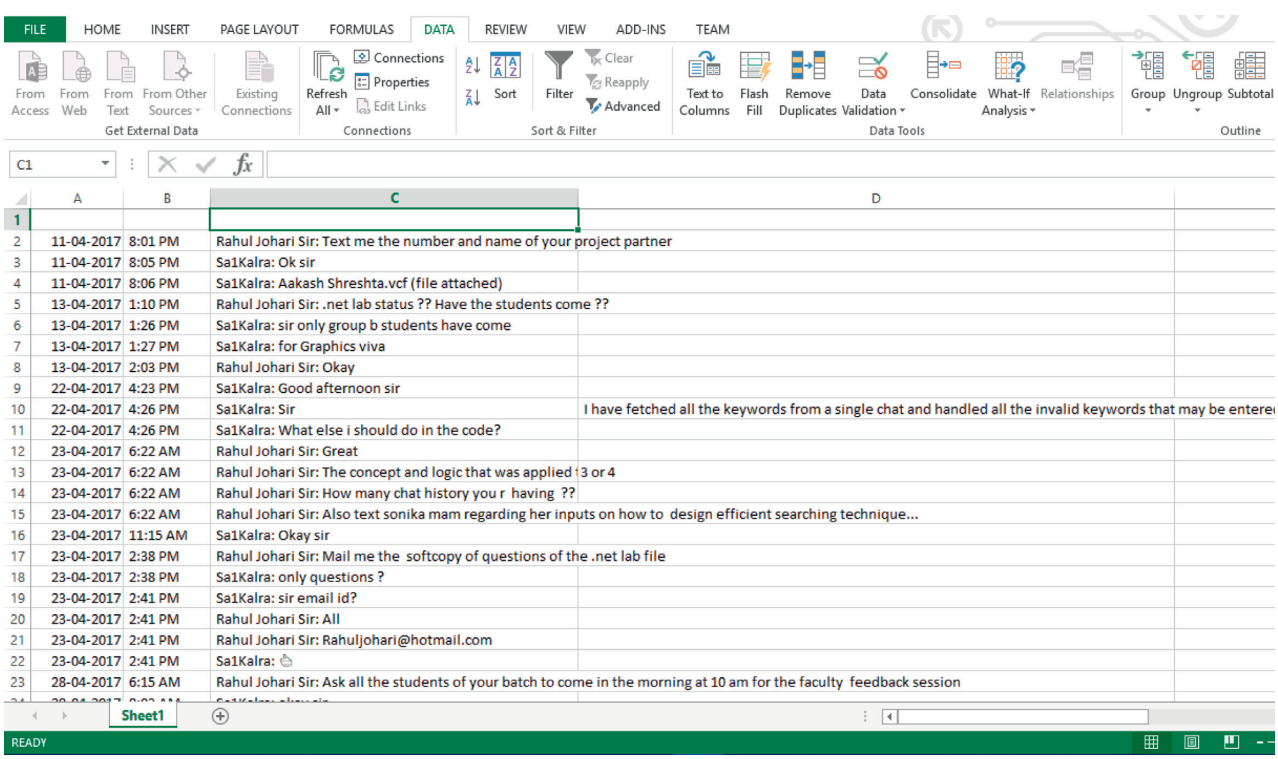
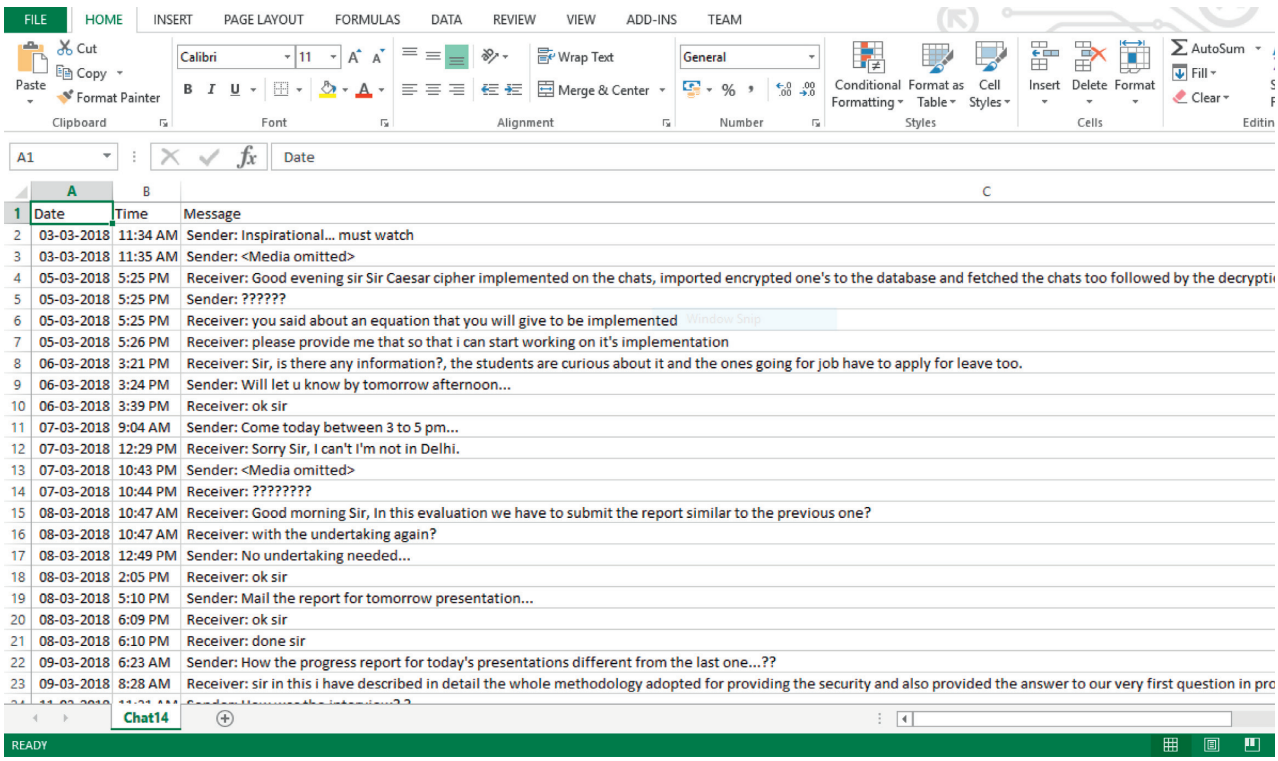


Figure 11: Text data separated into columns.



	A	B	C
1	Date	Time	Message
2	03-03-2018	11:34 AM	Sender: Inspirational... must watch
3	03-03-2018	11:35 AM	Sender: <Media omitted>
4	05-03-2018	5:25 PM	Receiver: Good evening sir Sir Caesar cipher implemented on the chats, imported encrypted one's to the database and fetched the chats too followed by the decrypti
5	05-03-2018	5:25 PM	Sender: ??????
6	05-03-2018	5:25 PM	Receiver: you said about an equation that you will give to be implemented
7	05-03-2018	5:26 PM	Receiver: please provide me that so that i can start working on it's implementation
8	06-03-2018	3:21 PM	Receiver: Sir, is there any information?, the students are curious about it and the ones going for job have to apply for leave too.
9	06-03-2018	3:24 PM	Sender: Will let u know by tomorrow afternoon...
10	06-03-2018	3:39 PM	Receiver: ok sir
11	07-03-2018	9:04 AM	Sender: Come today between 3 to 5 pm...
12	07-03-2018	12:29 PM	Receiver: Sorry Sir, I can't I'm not in Delhi.
13	07-03-2018	10:43 PM	Sender: <Media omitted>
14	07-03-2018	10:44 PM	Receiver: ????????
15	08-03-2018	10:47 AM	Receiver: Good morning Sir, In this evaluation we have to submit the report similar to the previous one?
16	08-03-2018	10:47 AM	Receiver: with the undertaking again?
17	08-03-2018	12:49 PM	Sender: No undertaking needed...
18	08-03-2018	2:05 PM	Receiver: ok sir
19	08-03-2018	5:10 PM	Sender: Mail the report for tomorrow presentation...
20	08-03-2018	6:09 PM	Receiver: ok sir
21	08-03-2018	6:10 PM	Receiver: done sir
22	09-03-2018	6:23 AM	Sender: How the progress report for today's presentations different from the last one...??
23	09-03-2018	8:28 AM	Receiver: sir in this i have described in detail the whole methodology adopted for providing the security and also provided the answer to our very first question in pro

Figure 12: Naming columns as date, time, and message.

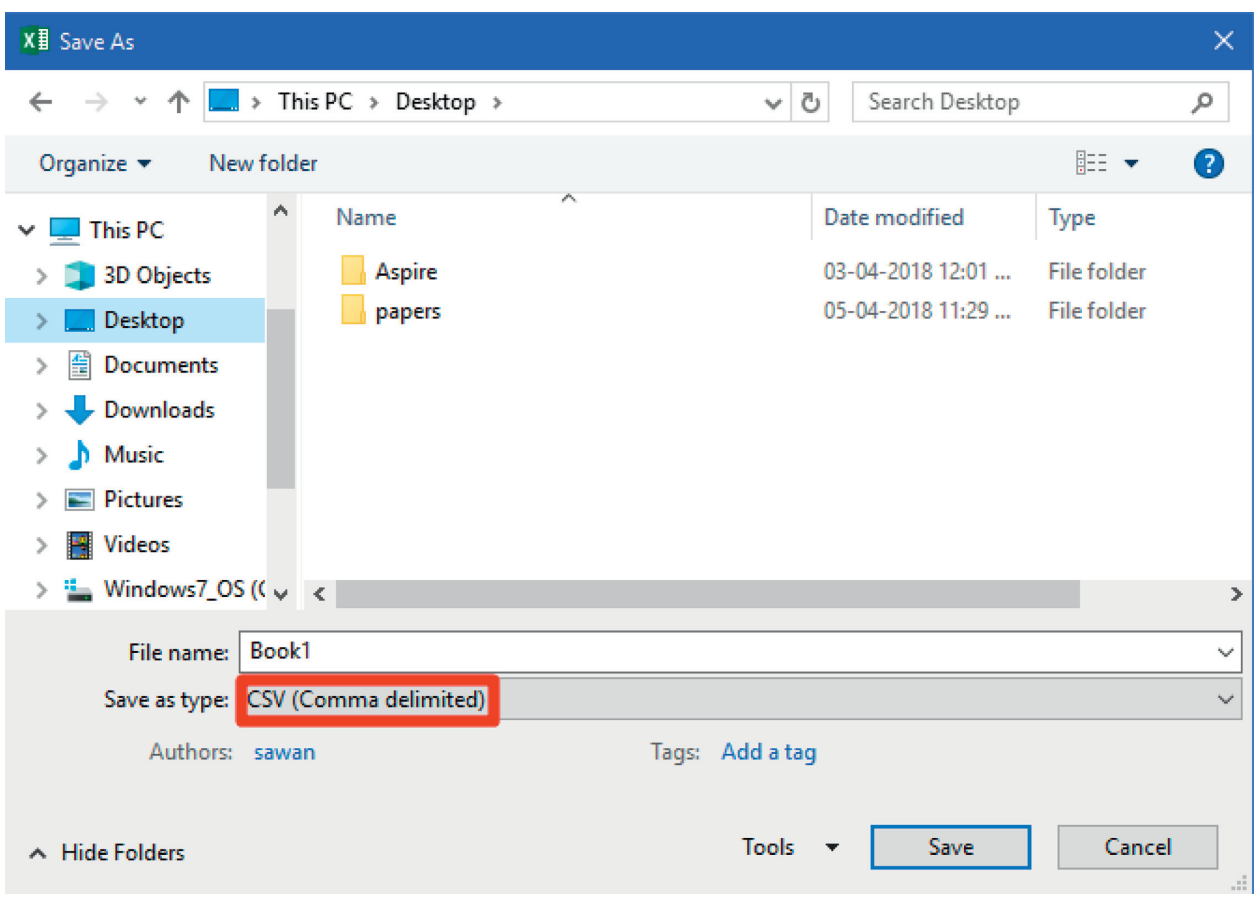


Figure 13: Saving of the spreadsheet in .CSV file format.

```
C:\WINDOWS\system32\cmd.exe - mongo
> use ceasar
switched to db ceasar
> db.chat1.find().pretty()
{
        "_id" : ObjectId("5ad8df198aacfa05fd2c1f5f"),
        "Date" : "01-03-2018",
        "Time" : "2:11 PM",
        "Message" : "Receiver:#?Phgld#rplwwhgA"
}
{
        "_id" : ObjectId("5ad8df198aacfa05fd2c1f60"),
        "Date" : "01-03-2018",
        "Time" : "2:11 PM",
        "Message" : "Receiver:#wkhhn#kdl#qd#ekdlBB"
}
{
        "_id" : ObjectId("5ad8df198aacfa05fd2c1f61"),
        "Date" : "01-03-2018",
        "Time" : "2:11 PM",
        "Message" : "Sender:#BBBBBB"
}
{
        "_id" : ObjectId("5ad8df198aacfa05fd2c1f62"),
        "Date" : "01-03-2018",
        "Time" : "2:12 PM",
        "Message" : "Sender:#\\h#ewd"
}
{
        "_id" : ObjectId("5ad8df198aacfa05fd2c1f63"),
        "Date" : "01-03-2018",
        "Time" : "2:12 PM",
        "Message" : "Sender:#Pmkh#hduskrqh#ohqh#k"
}
{
        "_id" : ObjectId("5ad8df198aacfa05fd2c1f64"),
        "Date" : "01-03-2018",
        "Time" : "2:12 PM",
        "Message" : "Receiver:#hydoxdwlrq#nd#pdw#sxfkqd#nxfk#edv"
```

Figure 14: Contents in the Caesar Cipher encrypted chat.

(5.4) Insert the words that have been selected into the respective dictionary with the number of occurrences of each and every unique word

(6) Repeat step 5 for left over fifteen WhatsApp chats numbered 2 to 16

(7) Create a complete connected weighted graph with 16 vertices, where each vertex represents a particular dictionary, and set the weight of edges to be zero

(8) At the end of step 6, it would be as in Figure 16

Now, 16 dictionaries were obtained, and one for each individual contains at least 10 words which are unique and whose length is greater than 6 along with the number of occurrences of that word. Out of 16 dictionaries, four random dictionaries are shown in step 4.

(9) From 1st dictionary, select the first word and search for its occurrence in dictionary 02 to 16

(10) If occurrence of the word is found in dictionary 02 to 16, then

increase weight by 1 of the edge between node representing dictionary 1 and node representing respective dictionary, where the word is found,

else

repeat step 9 and step 10 with the next word from dictionary 1 until all words in dictionary 1 are considered

Figure 15: Contents in the Block Quadra Crypto encrypted chat.

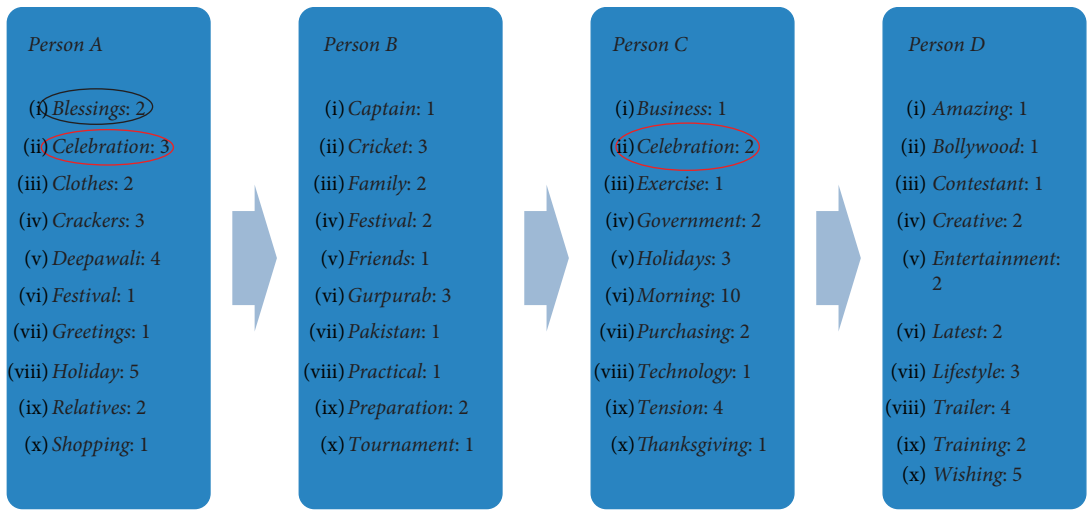


Figure 16: Words into the respective dictionary with the number of occurrences of each and every unique word.

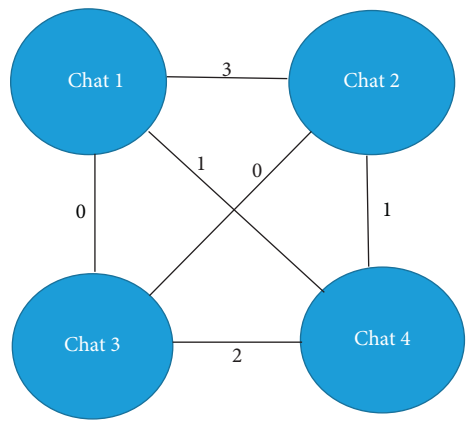


Figure 17: Sample graph to show the connection and linkages amongst chats.

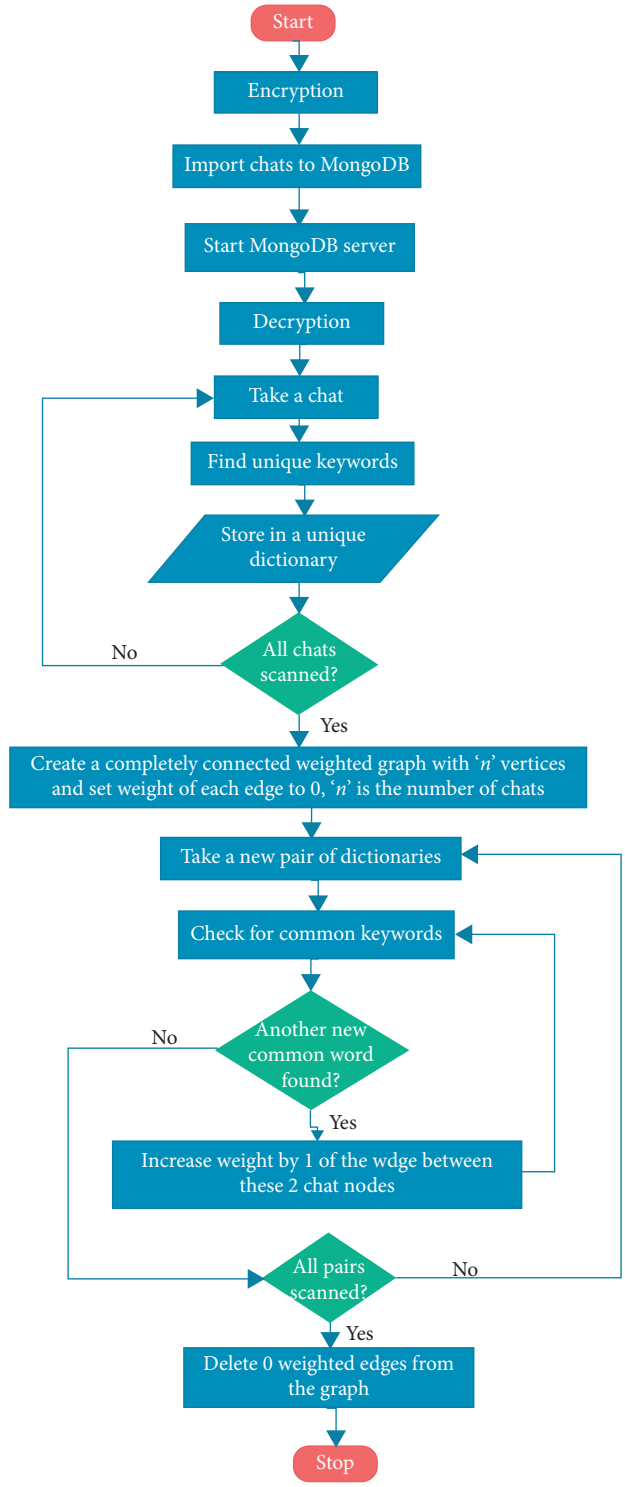


Figure 18: Flowchart indicating methodical procedure adopted for the efficient working of the S^2NOW algorithm.

(11) Repeat step 9 and step 10 for other dictionaries one by one, that is, from dictionary 02 to dictionary 16

(12) Delete zero weight edges

To conclude, primary focus is to detect social pattern amongst the selected sample dataset of 16 (sixteen) WhatsApp chats and it can amicably be concluded that greater the weight of edge, stronger is the bond.

6.2. Flowchart. Working steps of the S^2NOW algorithm, Secure Social Network Ontology using WhatsApp, are indicated in the flowchart depicted in Figure 18.

6.3. Mathematical Modeling. Notations:

WAC_i: WhatsApp chat of i^{th} person for a period of approximately 30 days/1 month

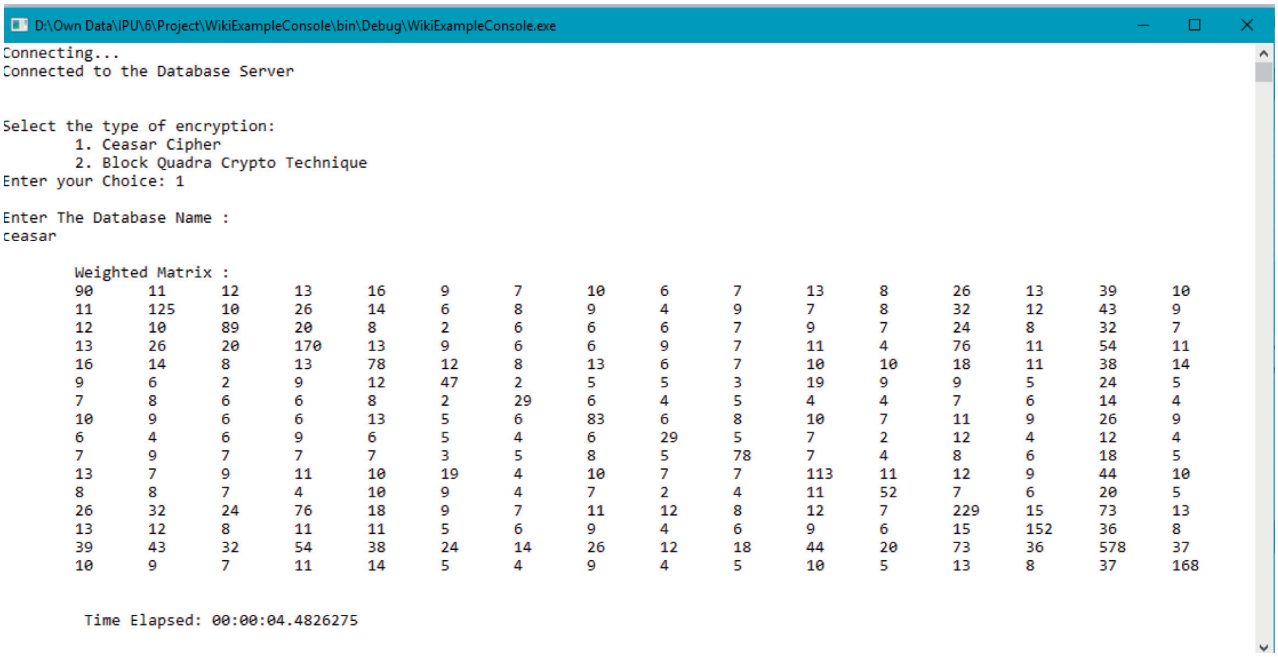


Figure 19: Resultant Weighted Matrix obtained on chats encrypted in Caesar Cipher.

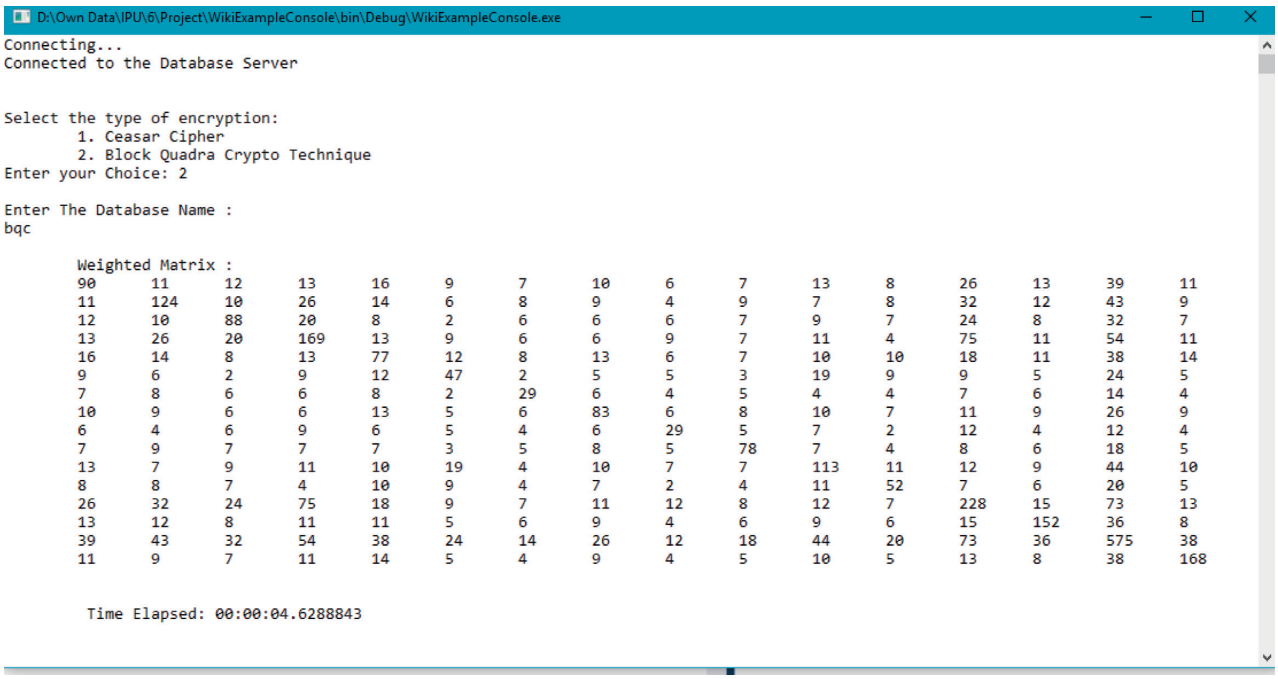


Figure 20: Resultant Weighted Matrix obtained on chats encrypted in Block Quadra Crypto Technique.

n: number of WhatsApp chats, one per person

Dict_i: a dictionary belonging to i^{th} WhatsApp chat, used for storing unique words

w_d: word in any WAC_i

l_w: length of w_d

INPUT: WhatsApp chats for n persons (WAC_i)

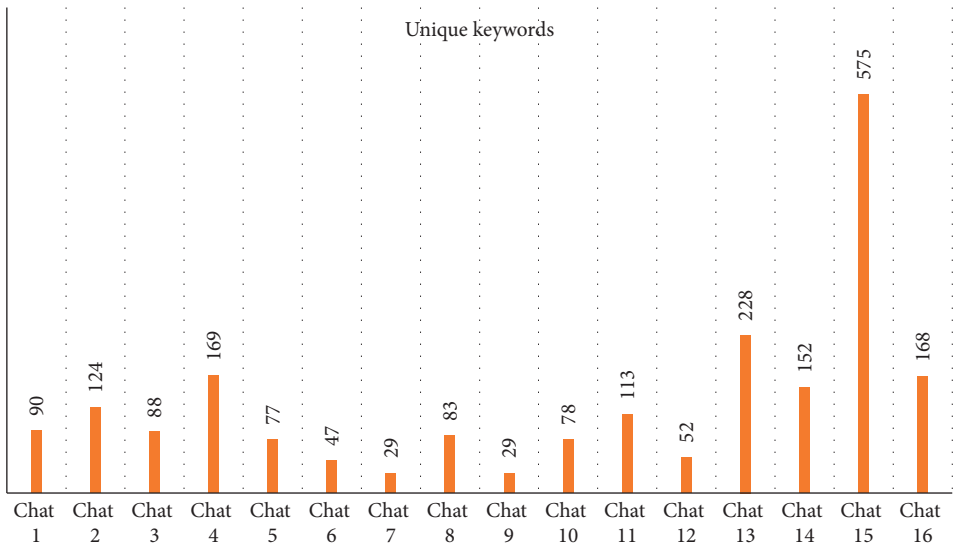


Figure 21: Bar chart showing the number of unique keywords found in the WhatsApp Chat of 16 persons.

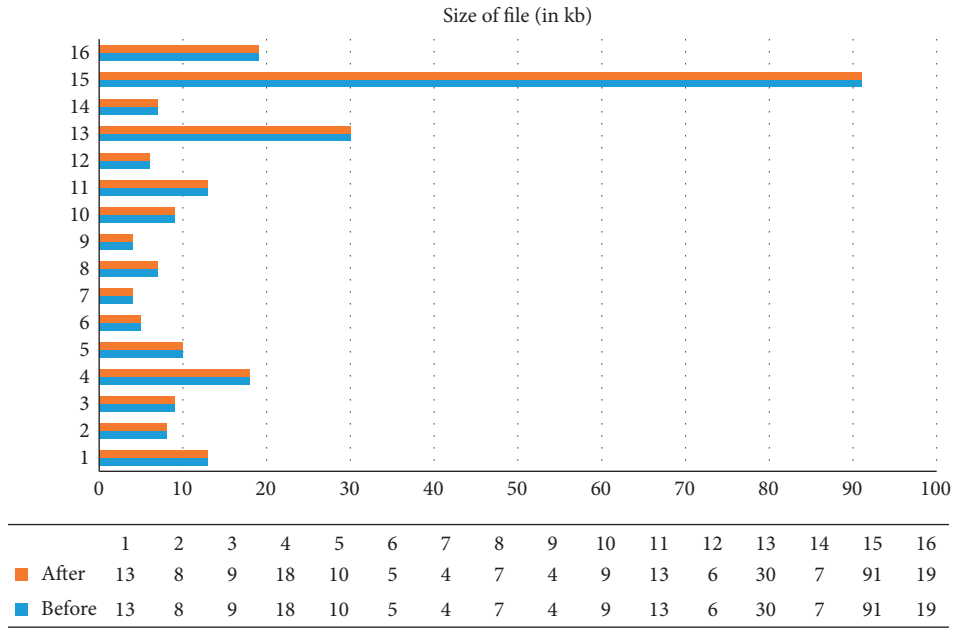


Figure 22: Bar chart showing the size of file before and after the application of Caesar Cipher.

OUTPUT: Weighted Graph with n vertices (G=(V,E))

(1) Set all Dict_i ϕ. $1 \le i \le n$

(2) Collect Sample WhatsApp chats of n persons for a defined date and time, say of a month

(3) Create weighted connected graph, G with $||V|| = n$. Set weight of each edge to zero

(4) **for** $i \leftarrow 1$ **to** n

do Start scanning from first letter of WAC_i.

for each wd $\in \text{WAC}_i$

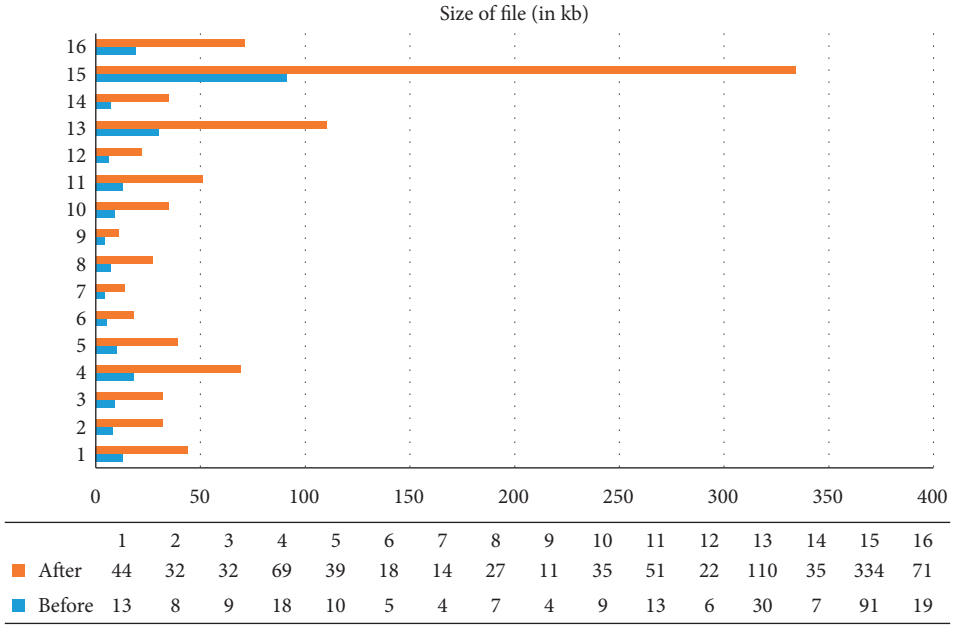


Figure 23: Bar chart showing the size of file before and after the application of Block Quadra Crypto Technique.

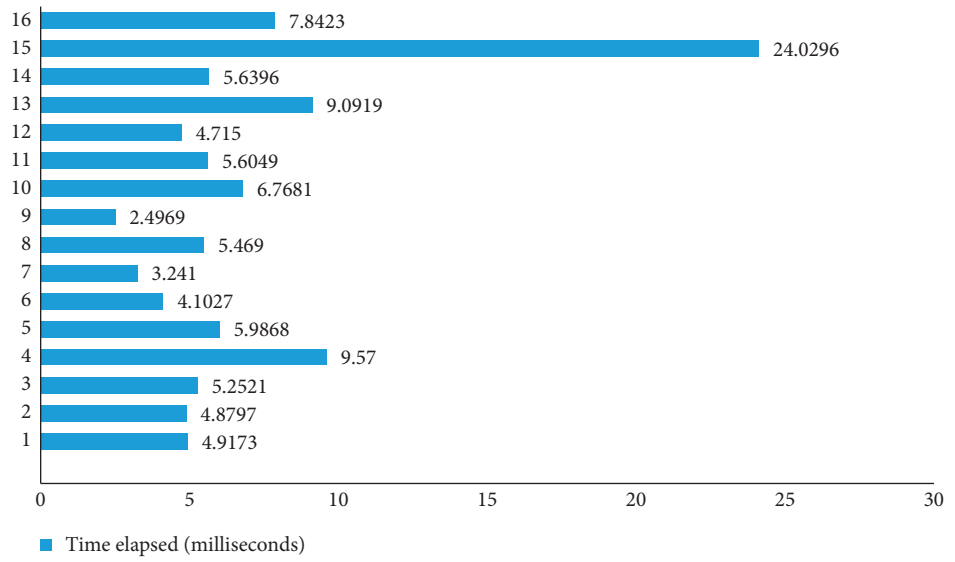


Figure 24: Bar chart showing the time consumed in encryption (Caesar Cipher).

do if ($l_w \geq 6$ and $w_d \in Dict_i$)
then continue
else if ($l_w \geq 6$ and $w_d \notin Dict_i$)
then $Dict_i \leftarrow w_d \cup Dict_i$

(5) **for** $i \leftarrow 1$ **to** n
do $j \leftarrow i + 1$
while $j < n$
do Check for each w_d from $Dict_i$
if (FIND(w_d, $Dict_j$))
do Increment weight of the edge between $Node_i$(representing $Dict_i$) and $Node_j$ (representing $Dict_j$) by one.

(6) Delete zero weight edges from the weighted graph

*FIND(w_d, $Dict_j$) returns true if $w_d \in Dict_j$ else false

7. Results

Social Network Resultant Weighted Matrix was created (Figure 19 with chats in Caesar Cipher and Figure 20 with chats in Block Quadra Crypto Technique) which shows the number of unique common words between 2 chats. For example, matrix position [1, 2] gives the number of unique common words in Chat 1 and Chat 2. The principle diagonal elements in the resultant weighted matrix represent the number of unique keywords in that chat; for instance, matrix position [2, 2] gives the number of unique keywords in Chat

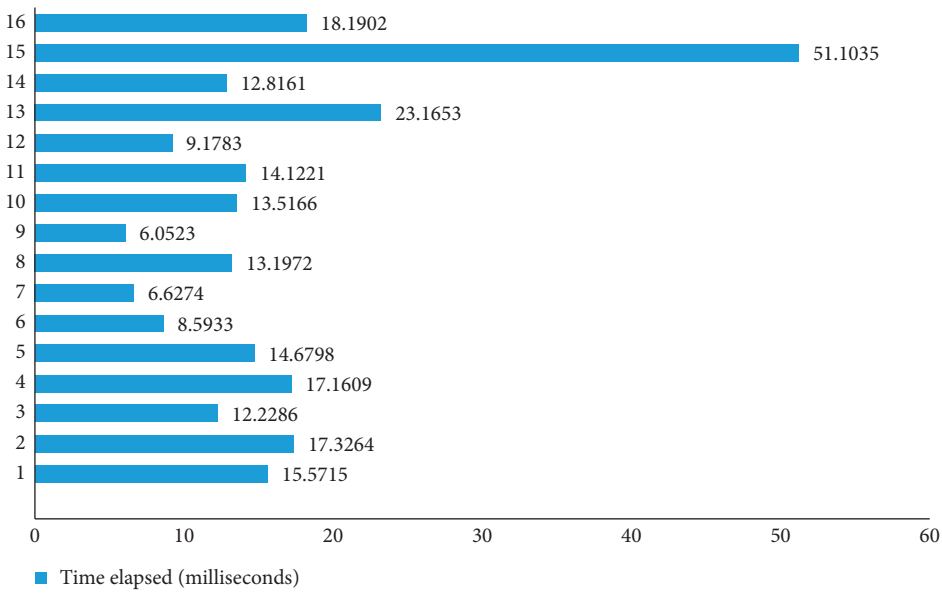


Figure 25: Bar chart showing the time consumed in encryption (Block Quadra Crypto Technique).

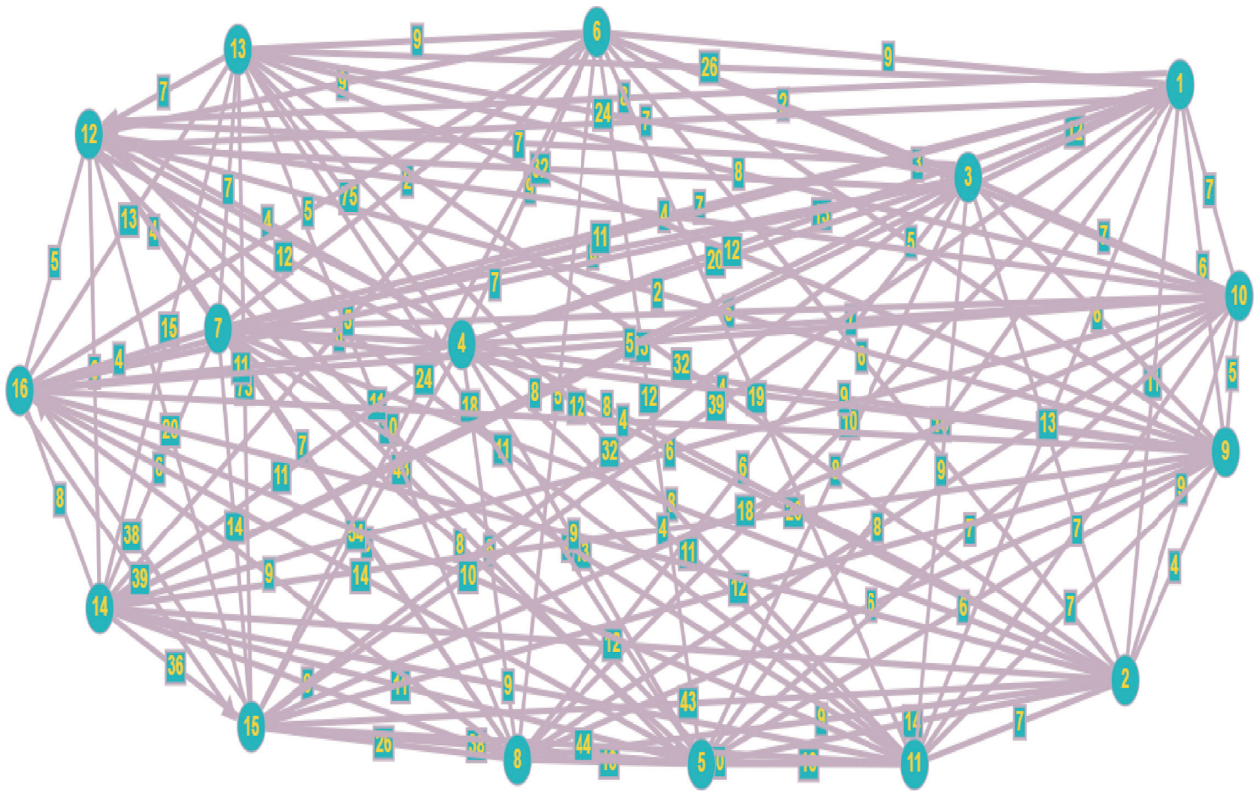


Figure 26: Node graph showing common words between 16 persons.

2. Figure 21 shows the total number of unique keywords found in the respective chats. Figures 22 and 23 show the change in the size of the .CSV (comma-separated value delimited) file before and after the application of the respective encryption technique; that is, the file size remains the same on the application of Caesar Cipher Technique, but

Figure 27: Node graph showing social inclination between groups of 16 selected persons.

on the other hand, the file size increases (approximately 4 times) on the application of Block Quadra Crypto Technique. Figures 24 and 25 show the time taken in milliseconds to encrypt the chat .csv files. Figures 26 and 27 depict some of the many ways of the node graph that can be represented where the nodes represent the persons and the edges represent the weights.

8. Conclusion and Future Work

This proposed work is in continuation of the previous work done in "WAPiS, WhatsApp Pattern Identification Algorithm indicating Social Connection" [24] wherein all the chats were imported into the MongoDB database and a connectivity program in C# was made.

In the current research work, the WhatsApp chats after encryption were imported into the MongoDB database and were then decrypted, before analyzing and the results were obtained. The implementation of cipher techniques was achieved, and the results as elaborated in Section 7 were achieved which are encouraging and exceed the expectations. However, it would be interesting to explore and enhance the asymmetric encryption cryptographic techniques and algorithms that facilitate the usage of the public and private key pair and would definitely provide a pretty large scope to work upon in the time to come.

In future, it also proposed to integrate the concept of community creation, identification, and detection in the current research work of collection and aggregation of the WhatsApp chats of different persons. This work would give insight into Ego Network and help in the quick identification of those nodes/persons who are quite active in the sharing and exchange of WhatsApp chats, popularly referred to as "Hub Nodes" in social network.

The results achieved are also proposed to be compared with other cryptographic techniques such as SCLCT, secured cross language cipher technique [25], C^3T, cloud-based cyclic cryptographic technique [26], and LBCLCT, location-based cross language cipher technique [27]. It is also proposed to extend the Block Quadra Crypto Technique with its successor symmetric cipher techniques such as Tri-Quadrant Cryptographic Technique (TQCT), Odd-Even Block-Based Cryptographic Technique (OEBBCT), and Quad-Quadrant Cryptographic Technique (QQCT). Also, effort would be made to enhance the level of security by adding a pair of public and private keys in the proposed cipher technique.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors are thankful to Director (Research and Consultancy), Guru Gobind Singh Indraprastha University, Dwarka, Delhi, India, for providing APC (article-processing charges) towards publication of this research paper.

References

[1] S. Bajaj and R. Johari, "Big data: a boon or bane-the big question," in *Proceedings of the 2016 Second International Conference on Computational Intelligence & Communication Technology (CICT)*, pp. 106–110, IEEE, Ghaziabad, India, February 2016.

[2] X. Chen, M. Vorvoreanu, and K. Madhavan, "Mining social media data for understanding students' learning experiences," *IEEE Transactions on Learning Technologies*, vol. 7, no. 3, pp. 246–259, 2014.

[3] S. Gole and B. Tidke, "A survey of big data in social media using data mining techniques," in *Proceedings of the 2015 International Conference on Advanced Computing and Communication Systems*, January 2015.

[4] N. Z. Quazilbash, S. M. H. Qadri, and S. Khoja, "Improved user RTSE experience on the web through fast retrieval of social media content," in *Proceedings of the 2012 15th International Multitopic Conference (INMIC)*, December 2012.

[5] Y. S. Tawiah, H. E. Nondzor, and A. Alhaji, "Usage of WhatsApp and voice calls (phone call): preference of polytechnic students in Ghana," *Science Journal of Business and Management*, vol. 2, no. 4, pp. 103–108, 2014.

[6] A. Gosain and S. Dahiya, "Performance analysis of various fuzzy clustering algorithms: a review," *Procedia Computer Science*, vol. 79, pp. 100–111, 2016.

[7] B. I. A. Alsaleem, "The effect of "WhatsApp" electronic dialogue journaling on improving writing vocabulary word choice and voice of EFL undergraduate Saudi students," *Arab World English Journal*, vol. 4, no. 3, pp. 213–225, 2013.

[8] J. Yeboah and G. D. Ewur, "The impact of WhatsApp messenger usage on students performance in Tertiary Institutions in Ghana," *Journal of Education and Practice*, vol. 5, no. 6, pp. 157–164, 2014.

[9] A. Ngaleka and W. Uys, "M-learning with WhatsApp: a conversation analysis," in *Proceedings of the 8th International Conference on e-Learning ICEL-2013*, p. 282, Academic Conferences International Limited, Cape Town, South Africa, June 2013.

[10] C. Barhoumi, "The effectiveness of WhatsApp mobile learning activities guided by activity theory on students' knowledge management," *Contemporary Educational Technology*, vol. 6, no. 3, pp. 221–238, 2015.

[11] T. Bansal and D. Joshi, "A study of students experiences of WhatsApp mobile learning," *Global Journal of Human-Social Science Research*, vol. 14, no. 4, 2014.

[12] D. Bouhnik and M. Deshen, "WhatsApp goes to school: mobile instant messaging between teachers and students," *Journal of Information Technology Education: Research*, vol. 13, pp. 217–231, 2014.

[13] A. Sánchez-Moya and O. Cruz-Moya, "WhatsApp, textese, and moral panics: discourse features and habits across two generations," *Procedia - Social and Behavioral Sciences*, vol. 173, pp. 300–306, 2015.

[14] A. D. Ahad and S. M. A. Lim, "Convenience or nuisance?: the 'WhatsApp' dilemma," *Procedia - Social and Behavioral Sciences*, vol. 155, pp. 189–196, 2014.

[15] C. Petitjean and E. Morel, ""Hahaha": laughter as a resource to manage WhatsApp conversations," *Journal of Pragmatics*, vol. 110, pp. 1–19, 2017.

[16] T. Simon, A. Goldberg, D. Leykin, and B. Adini, "Kidnapping WhatsApp - rumors during the search and rescue operation of three kidnapped youth," *Computers in Human Behavior*, vol. 64, pp. 183–190, 2016.

[17] https://www.darkreading.com/endpoint/whatsapp-mobile-phishings-newest-%20attack-target/a/d-id/1332652.

[18] https://www.cybersecurity-insiders.com/millions-of-whatsapp-and-telegram-%20users-are-vulnerable-to-cyber-attacks/.

[19] https://www.cybersecurity-insiders.com/cyber-attack-on-whatsapp-leads-to-a-%202-hour-major-outage/.

[20] https://research.checkpoint.com/fakesapp-a-vulnerability-in-whatsapp/.

[21] https://searchsecurity.techtarget.com/news/252446722/WhatsApp-%20vulnerabilities-let-hackers-alter-messages.

[22] https://searchsecurity.techtarget.com/answer/How-did-WhatsApp-%20vulnerabilities-get-around-encryption.

[23] https://www.bleepingcomputer.com/news/security/whatsapp-vulnerability-%20allows-attackers-to-alter-messages-in-chats/.

[24] R. Johari, S. Kalra, S. Dahiya, and P. Yadav, "WAPiS: WhatsApp pattern identification algorithm indicating social connection," in *Proceedings of the International Conference on Advanced Computational and Communication Paradigms – 2017 (ICACCP-2017)*, pp. 595–603, Springer, Sikkim, India, September 2017.

[25] S. Kumar, R. Johari, L. Singh, and K. Gupta, "SCLCT: secured cross language cipher technique," in *Proceedings of the 2017 International Conference on Computing, Communication and Automation (ICCCA)*, pp. 545–550, IEEE, Greater Noida, India, May 2017.

[26] S. Gupta, R. Johari, P. Garg, and K. Gupta, "C^3T: cloud based cyclic cryptographic technique and it's comparative analysis with classical cipher techniques," in *Proceedings of the 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 332–337, IEEE, Noida, India, February 2018.

[27] V. Gupta, R. Johari, and K. Gupta, S. Seth, LBCLCT: location based cross language cipher technique," in *Smart Cities Performability, Cognition, & Security*, pp. 221–234, Springer, Switzerland, 2020.