Semantic Network Analysis of Recent Twitter Security

Mansour Alshaikhsaleh
School of Engineering and Computer Science
Oakland University
Rochester, MI, United States
Email: Mansouralshaikh [AT] oakland.edu

Mohamed A. Zohdy
School of Engineering and Computer Science
Oakland University
Rochester, MI, United States

Abstract—**As the technologies advance rapidly, the privacy and the security have become one of top concern in the digital communication in this age. Thus, we are trying to understand and drawing a conceptual view depend on how is the people viewed privacy and security in this era which has been changed and still changing and expanding as the technology developed. In order to analysis and see what it the person's apprehension regarding the view of the privacy and security, we have used one of the widespread social media which is Twitter. Based on the Twitter's users view we have constructed our research and draw our conclusions. We have collected more than 20,000 tweets to analysis them based on different aspects. Our results discovered a number of significant implications for privacy and security from the tweet users' impression. In short, the mobile security, Facebook security hole and the privacy of the account information, privacy of the mobile data and who could access it. These are the top concerns in these days regarding the security and privacy from the Twitter's user's perspective.**

Privacy; Security; Twitter Analysis; Semantic Network Analysis; Social Media

I. Introduction

This project mainly focused on privacy and security then it takes into account the Health Information Portability and Accountability Act (HIPAA), iCloud, information security and hats privacy –white, gray and black hat privacy- key words. After collecting the tweets, we have analyzed them using AutoMap tool. In the initial analysis, AutoMap extract all of the words which has been occurring in the tweets along with each word frequency. As a second step of the analysis, we have filtered out all about numbers, symbols, and non-relative wording. Then, we took the most frequent work and did discourse analysis to see how these words are emerging. Based on discourse analysis, we have wiped out some of the word that appears in different context. From that point we have drawn our discussion and conclusions.

Based on our analysis, the privacy is beyond the information control and access. The people have changed their view regarding the privacy with the advancement of the technology. The Twitter's users get concerned about their social media personal data privacy. Particularly, when a company change their privacy terms such as, the Facebook has posted the new term and privacy policy, the users become more worried about their privacy. In addition, Twitter users were discussing the mobile devices privacy and security to further assist the people who plan to buy in the near further. Moreover, they have anxiety regarding different acts and lows that are developed, and whether these laws will protect them or the laws will work against them. When the security is present with the privacy, we see the trend of talking about the cloud, HTC, Cyber Intelligence Sharing and Protection Act (CISPA), surveillance, and Health Information Portability and Accountability Act (HIPAA). Even in this context, the people classified different technology based on their robustness like comparing Android or Apple devices.

When the users talking about HIPAA, we found how the people feel more secure with the new added rules regarding the patient records. Beside other issues which has happened, such as the privacy breaches and leaks in the health system. On the other hand, Twitter's users further discussed how they can protect their personal information and hide it. Adding to the preceding concerns, the online security is one of top topic in the security and privacy context; thus, Twitter's users provide advices on how to write a robust password.

A. *Statement of the Problem*

As the new generation moving toward technology, the social media has played anrole in reflecting the people ideas and thoughts without any bias [1]. Understanding the society views of specific topic or phenomena involves a sophisticated cognitive and planning process. In the context of privacy and security, we would like to have an impression of how different people from different place relate the security and privacy with other contexts. Thus, it will lead us to produce a satisfactory software or experience with these people. One way to understand the people sight on relating the security and privacy terms with other terminologies is studying their ideas on the social media like Twitter and Facebook [1]. We can extract valuable information from different social media by analyzing their data [2]; therefore, by analyzing the posted tweets, we can have an overview picture of how Twitter users discuss the relations between the security and privacy thus contributing to how the semantic networks of the two terms develop. In this project, we are trying to visualize how the Twitter users link the security and privacy terms with other keywords. This study examines the discourse around privacy and security on Twitter. It does so through semantic network analysis and discourse analysis of tweets containing the words "privacy" and "security."

B. Definition of Terms

In the current era, with the movement to technology and Internet even the people social life; it became clear how the privacy and security turn out to be important factors to consider. In the beginning, we need to define the privacy and security terminologies. Defining privacy is challenging, as it is a very elastic concept. Privacy is a social process whose management depends on people's interaction and social exchange as well as physical nature of the interaction. In addition, manifestations of privacy are culturally sensitive and contextual based. On the other hand, if we define the security in the computer context, we have to highlight three important characteristics of any computer system which are confidentiality, integrity, and availability [3].

II. Methodology

A. Data Collection

This part has been done by searching the tweets which contain privacy and security key word using an online application called HootSuite Archives:

- The first set includes all the tweets that have the keyword "security".
- The second set includes all the tweets that have the keyword "privacy".
- The third set includes all the tweets that have both keywords "privacy" and "security".
- The fourth set includes more specific tweets which have both keywords "information security".
- The fifth set includes all the tweets that have the following keywords "HIPAASecurity".
- The sixth set includes all the tweets that have that have the following keywords "icloud privacy".
- The seventh set includes all the tweets that have that have the following keywords "icloud security".
- The last three sets include all the tweets that have that have the following keywords "white hat privacy", "gray hat privacy" or "black hat privacy".

B. Data Analysis

We have used AutoMap to create a concept list which a list of each word and its frequency of occurrence in the collected tweets.

C. Data Visualization

Visualization is our final step which we have used a word cloud tool to visualize our finding and how the Twitter users' relate different keywords to the security and privacy terms.

D. Research Parts Categorizations

We have two major parts in our study which are:

1) Automated Part

In particular, using semantic network analysis software such as AutoMap [4] and WORDij [5], we have analyzed them to draw a conceptual map that shows what other words are used in relation to privacy and security. The visualization of the networks implemented using available online software after we implement a function to clone the word based on its frequency. The quantitative analysis will include two steps. First, we investigated the semantic network of tweets by including every word of the tweet in the analysis. This step provided a more comprehensive, yet less specific, conceptual map of privacy and security that emerge from users' discourse. Second, we analyzed the keywords that users have intentionally connected to privacy and security using the symbol "#." The "#," or "hashtag," is used on Twitter to identify keywords and key-phrases. This step will provide a semantic network of the core elements that users purposely relate to privacy and security. Emerged semantic networks further our understanding of users' view of the intersections between privacy, security, and other "keywords." Keywords, for example, may include legislations as the Health Insurance Portability and Accountability Act (HIPAA) and the Family Educational Rights and Privacy Act (FERPA). Or they may include other key-concepts that often emerge in relation to privacy and security, as "trust" (i.e. [6]) or "surveillance" (i.e. [7], [8]).

2) Manual Part

Finally, we randomly selected a sub-sample of tweets to be further analyzed through discourse analysis. As semantic network analysis is an automated process and its results missed some nuances of the discourse brought about in the tweets analyzed. Thus, the text needs to be further examined using a more in-depth approach. Discourse analysis is a qualitative process seeking to provide a deeper explanation of meaning through the analysis of themes and patterns that emerge from texts. It also takes into account the role of context in developing the semantic networks of "privacy" and "security." Such a qualitative approach strengthens the findings obtained in the quantitative steps of this research project.

III. Significance of the Study and Potential Outcomes

Even so there is a lot of literature that investigates aspects of privacy and security, not much has been done to attempt understanding how the concepts of privacy and security are perceived and communicated among individuals, and how they intersect with other important constructs. Yet this is a fundamental element to investigate. The literature is rife with mentions of "privacy" and "security", but often fails to define them. By investigating how the concepts of privacy and security are discussed and individually framed, this project may contribute to drawing a semantic map of the terms. The outcome of this analysis will be highly valuable to any private or public institution interested in the interconnection between privacy and security and in how such interconnection develops in the mind of users. In addition, the results of the analysis could assist in the security improvement on different apps and online networks etc.

IV. Literature Review

A. *Theories of Privacy*

In modern Western societies, privacy is often recognized as a basic condition for individual autonomy, liberty, identity, and integrity ([9], [10], [11], [8], [12]). The concept of privacy has been thoroughly explored in several disciplines. Yet, despite increased attention, scholars have not reached an agreement on how privacy may be theoretically or empirically defined. The boundaries of privacy are fluid and undergo incessant renegotiations that, partly, depend upon the rapid advances in information and communication technology and the increased capacity to collect and process personal data. Scholars have broken up and analyzed the notion of privacy but there has been no complete uniformity on its definition. Privacy viewed as a dynamic process of access control, and a fundamental means to personal autonomy, emotional release, self-evaluation, and limited and protected communication. researchers framed privacy focusing on matters of accessibility and boundaries control, and suggesting that dissonance emerge when desired and actual levels of privacy differ. Building on the previous view dialectic understanding of privacy, Child and Starcher [13] explained the process of boundaries negotiation through the communication management theory (CMT). CMT assumes that individuals own personal information and, thus, develop privacy rules to regulate its flow. Cultural values, social norms, contextual impact, and cost-benefit analysis intervene in the individual designation of privacy rules [13].

In addition, privacy emerges is a bi-directional process of access negotiation that operates at the individual and at the group level. Thus, privacy may be intended as an individual value, a societal value [14], and a cultural universal [15].

Research has investigated the multifaceted nature of privacy and its relationship to other values, providing a number of fundamental perspectives to further understanding the complexity of privacy (e.g. [9], [16]). When discussing values, philosophers traditionally distinguish between instrumental values, which are means to achieve some other ends, and intrinsic values, which are self-justifying ends. A common approach suggests that privacy has instrumental value, as it is necessary to protect one from specific practical harms ([17], [18]). To claim that privacy has higher instrumental values, one should prove the role of privacy in achieving higher-level benefits. Stahl [12], for example, argued that privacy is necessary to maintain a variety of relationships. Suggesting the intrinsic value of privacy, some researchers explained that privacy is a fundamental aspect of autonomy. Similarly, Taddeo [19] acknowledged that privacy is not a core value per se. Yet, he argued that when a society reaches a certain level of development and becomes less intimate, privacy evolves into security needs. Privacy, in other words, protects us from strangers. Each person should be considered as constituted by his or her information; in such a view, privacy becomes pivotal for one's identity. Adopting this perspective, one may view personal information as one's substance or essence, rather than as one's property. With no privacy, one cannot achieve personal wellbeing or maintain personal identity. Theories of privacy have approached it from normative perspectives ([10], [20]) or adopting merely descriptive approaches [21].

Overall, discussions around privacy have entailed a variety of dimensions that include psychological, legal, social, and informational elements [22]. To address privacy, Nissenbaum in [11] developed the framework of contextual integrity to explain that privacy is respected when information about the self is as private as one assumes it to be. Addressing the philosophical, legal, economic, and cultural debates around privacy, Nissenbaum in [11] provided a valuable foundation to understand the intersection of privacy, electronic security, surveillance, affordances, and individuals' expectations in the digital environment. In particular, the framework of contextual integrity may embrace the complexity of privacy as it recognizes that privacy is an elastic concept whose meanings evolve and are renegotiated across contexts. Such a framework takes into account norms, social roles, information types, and transmission principles attached to information. Thereby, it may allow one to address the complex relationship of privacy and security in different domains. Scholars have identified different concepts of privacy often connecting it to three fundamental aspects: the creation of knowledge, the value of dignity, and the freedom right ([17], [23]). The first meaning of privacy, related to the formation of knowledge, implies the sense of violation that originates when information about the self-leaks, potentially polluting one's image. Receivers may use that information to form judgments based on data wrenched out of context rather than on intimate knowledge [24]. Privacy, instead, enables one to manage the flow of information and avoid misrepresentation and distorted knowledge ([23]; [8]). The idea of privacy as the management of what other people know about the self may be exemplified suggesting that the collection of data about an individual creates one's portrait, "an evocative depiction, meant to convey something about the subject's character or role in society" [24]. Problems emerge when such a portrait differ from desired representations of the self [8]. Critiques to this approach suggest that the risk of public misconception may relate to privacy infringements, but it is not necessarily a matter of privacy. Adopting this meaning of privacy, one can begin to see the interconnections between privacy, sense of self, and sense of belonging that stem from perceived self-efficacy in the management of one's personal image and reputation.

The second meaning of privacy discusses its relation to individuals' dignity ([21]; [17], [23]). Dignity, which refers to one's sense of self as something worth respecting, may contribute to shaping one's identity. Scholars who have adopted this perspective have argued that privacy infringements may generate harm due to the violation of "significant normative expectations" and constitute an intrinsic offense against one's dignity. Such an offense is harmful as it hinders one's ability to save public face ([25], [8]). The tendency to associate privacy with dignity is a typical aspect of European privacy rights and laws that protect one's name, image, and reputation ensuring a shield against public exposure [25].

Finally, a third approach to privacy connects it to the value of liberal freedom suggesting that "a liberal state respects the distinction between public and private speech because it recognizes that the ability to expose in some contexts parts of our identity that we conceal in other contexts is indispensable to freedom" [23]. This sense of privacy stems from a concept of freedom that originated in the Renaissance or in the Reformation. The understanding of privacy as freedom entails the negotiation between social norms and self-interest, but focuses on the latter. Adopting the approach of placing the emphasis on the individual over the community, some have intended privacy as freedom from external social norms epitomized in the family. Others have interpreted it as a liberty from state regulation and from intrusion within the domestic walls ([25], [23]). The latter is often associated with the approach adopted in American laws. Considering our goal to relate privacy and security aspects, identity is not only holding a privacy matter, but also could lead to security breaches via identity theft [26].

Analyzing the value of liberty, researchers highlighted a fundamental difference between negative and positive freedom. On the one hand, negative freedom - or freedom from - entails opportunities and depends upon the lack of coercion due to external interference with one's activities.

On the other hand, positive freedom - or freedom to - derives from the individual's desire to be one's own master, and the ability to follow one's conscious purposes thereby self-determining one's success or failure. Natural freedom, if unlimited, would lead to social chaos. To avoid such a risk, the area and limits of free action must be determined by norms or laws. However, it is challenging to designate a fair balance between one's and others' freedom. Those who have identified the value of privacy with that of civil liberty have suggested that, for example, decisional privacy relates to one's freedom to choose [16].

B. Empirical Studies of Analyzing Twitter

Day after day, Twitter is hosting important information in which the Twitter users are willing to share; thus, new opportunities of extracting and analyzing information emerge. One of the researcher groups has taken the advantage of such availability. They analyzed the tweets using a combination of content and structural analysis approach. Using the social network analysis, they have identified the reliable and trustworthy tweets. Then they digested the "key emergency" evidences. The researchers built a novel framework to analyze Twitter and text on the publicly available tweets. They claimed that their framework is fast and process enormous tweet streams in real time [27].

Marcin and Shiu [28] studied Twitter and Wikipedia to extract the topic trends by correlation Twitter and Wiki topics. The researchers claim that Wikipedia is a brilliant source in this regard since each Wiki page is about one specific topic. They have grouped tweets according to specific time frames and represented them in semantics way using word distribution. The researchers also tried to find a relation between the topics and the time these topics were discussed and introduced. They were collecting the tweets over six-month period then validate their findings of the trends by comparing it with Google's search volume data. The researchers used semantic network analysis to connect and extract the topics from Twitter and Wikipedia Datasets. They successfully verified the correctness of their approaches by comparing their results with Google trends when using search volume data. Similarly, Zubiaga et al. [29] found the topic trends in real-time by studing Twitter. Parra et al. [30] considered Twitter usage in academic evant. There main focus was on 16 computer science conferances. Examining tweets on Twitter has gone beyond just analyzing to extract topics or knowing "key emergency", for example, it has even hit the business and marketing world [31]. Okazaki et al. [32] have analyzed tweets to investigate the customers' opinions on brands as an electronic word-of-mouth (WOM). In their research, they have analyzed tweets content, post time, range, and frequency. Furthermore, they have found the percentages of the positive, negative, and criticism of the product or the company posts. One of the research goals is to enhance and build the relationship between the customers and company since Twitter provides a platform in which the connection happens almost in real time.

In our literature review, most of the researchers are referring to Twitter as microblogging. Microblogging considered as a form of communication in which the users write short messages using instant messaging, emails, web, or mobile phones, etc. By this definition, Twitter considered one of microblogging widespread systems. The researchers have used network semantic analysis in microblogging community. Where they have studied Twitter as a social network, and observed the connection between the topological and geographical possessions. The researchers focus was on the users who used Twitter to share their daily activities, information, and status. Moreover, they are proving that the users who have similar intentions most likely to connect with each other [33]. They have used a two-level approach combining HITS algorithm [34] and community detection. HITS algorithm used "to find the hubs and authorities in the Twitter social network." The hubs and authorities share a reinforcing asset. The authority value is the sum of the user scaled hub values of his/her followers, and the user hub value is the sum of the scaled authority values of the users that she/he is following. After applying the HITS algorithm, they have detected the existing communities by identifying the friendship relations between the users; they have considered the bidirectional links only to assure that they are a friend and sharing some interests [33].

In this research, the researchers were using Twitter and analyzing selected tweets to detect phishing which is considered as a cybercrime. Phishers are people who try to steal users' personal information to use them for fake purposes. The research goal was to understand the mechanism of phishing in online social media. Since Twitter has limited space for text, phishers tend to use URL shortening services – makes the length of the URL shorter and hides the actual length of the URL behind it- for their purpose besides hiding their identity. In the research, the researchers were after the impact of phishing in online social media [35].

One of the papers described Twitter users view about the energy issues, frames and behaviors; unlike our research which is trying to capture the users' views on the privacy issues. Nevertheless, our methodology is similar. We start with collecting tweets then get the frequency, semantics, and context by analyzing the collected tweets as one of the most popular social media these days. The investigators in their research have shown an initial analysis of the tweets that have been collected in over 4 months. Their primary result of analyzing the tweets is the feasibility of awareness on the collected Twitter streams, beside defining the users' attitudes and gather inspiration in communication that encourage the people to change their behavior toward energy [36].

The researchers presented an overview of a state-of-the-art for analyzing and mining social network with different business techniques and applications. As mentioned before the social network has played an important role in the business and marketing world; thus, the companies' vision and strategy could be altered by social networks like Twitter. Even though collecting information from a social network could have different problem, for example, data duplication, inactive users, and automated agents, but the information which is available on these networks still valuable. Therefore, the paper's authors discussed different issues from the business applications perspective which is related to collecting data from social networks, such as data acquisition and preparation [31].

V. Results

The first step in conducting the project result is the quantitative analysis. Using Hootsuit, we have collected more than 10000 tweets –including the re-tweets- focused on "privacy". Figure 1 presents a bar graph of relative percentage for each word in the top 50 most frequent cross texts. This based on the collection of 10000 tweets that has the word "privacy" that has been collected over February 24 to March 14, 2016.

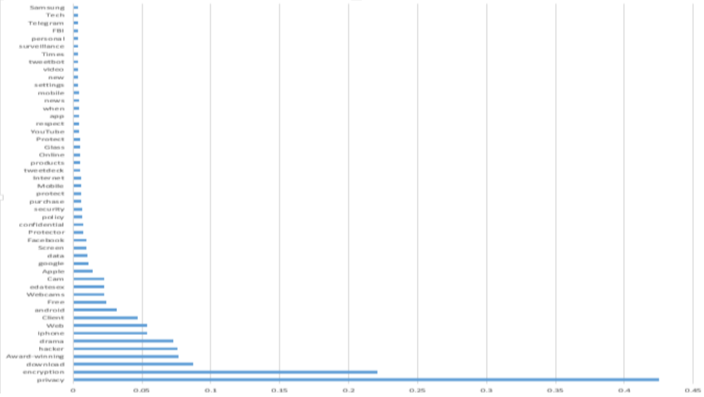

Figure 1. Relative percentage cross texts for the most frequent word occurs on the collection of 5000 tweets –without the re-tweets- that has the word "privacy".

Figure 2 represent the words in cloud based on relative most frequent words in privacy collection.

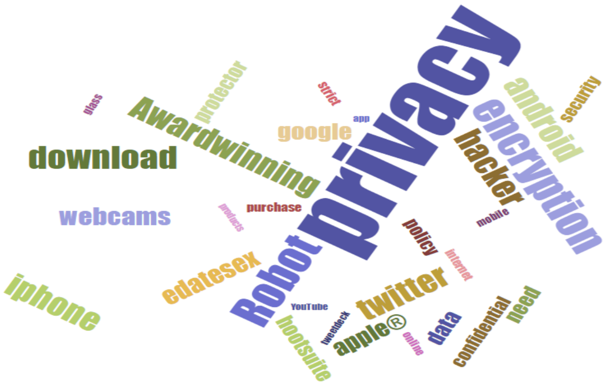

Figure 2. Cloud representation of the relative most frequent words that appears in the "privacy" dataset.

Table I includes a list of top 30 most frequent words as a result of the frequency of analyzing the collected tweets that include the word privacy after the tweets has been filtered and from around 10000 words, we studied less than 500 words.

TABLE I. THE MOST FREQUENT WORD OCCURS ON THE COLLECTION OF 5000 TWEETS –WITHOUT THE RETWEETS- THAT HAS THE WORD "PRIVACY".

Sequence	Concept	Frequency
1	privacy	5143
2	Robot	2117
3	encryption	1987
4	download	1384
5	twitter	1360
6	Award-winning	1358
7	hacker	1348
8	iphone	1343
9	android	1341
10	webcams	931
11	edatesex	970
12	apple®	850
13	google	843
14	data	752
15	need	691

16	hootsuite	674
17	protector	629
18	confidential	529
19	policy	519
20	security	458
21	purchase	213
22	strict	197
23	internet	151
24	mobile	104
25	glass	96
26	online	76
27	products	68
28	tweetdeck	51
29	YouTube	40
30	app	31

To look closely, the second step in steering the project result is to see the connection between the most frequently occurring word when the Twitter users talking about the privacy and security in the same context and when they talk about the privacy regardless if they talk about the security or not. Using the same tool, we have collected more than 9000 tweets focused on "privacy" and "security" in conjunction.

TABLE II. THE MOST FREQUENT WORD OCCURS ON THE COLLECTION OF 5200 TWEETS –WITHOUT THE RETWEETS- THAT HAS THE WORD "PRIVACY" AND "SECURITY".

Sequence	**Concept**	**Frequency**
1	privacy	5638
2	security	5405
3	twitter	3508
4	twitterfeed	3251
5	amp	3026
6	hootsuite	2497
7	data	2132
8	iphone	2075
9	download	2056
10	tweetdeck	1277
11	android	1051
12	dlvr	920
13	tweet	894
14	mobile	856
15	htc	851
16	tweetbutton	778
17	button	776
18	facebook	596
19	google	572
20	internet	260
21	news	117
22	https	104
23	surveillance	100
24	apple	99
25	hipaa	87
26	infosec	80
27	cloud	78
28	protection	57
29	cispa	28
30	online	15

Table II presents a list of top 30 most frequent words as a result of the frequency results of analyzing the collected tweets that include the word "privacy" and "security" after the tweets has been filtered as we did for the above collection. Figure 3 represent the relative words in cloud based on their frequency.

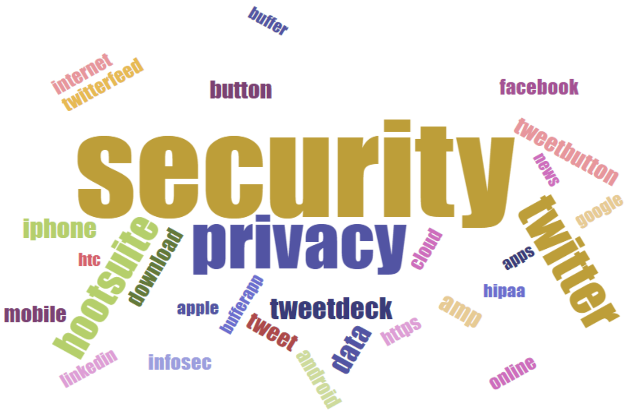

Figure 3. Cloud representation of the relative most frequent words that appears in the "privacy" and "security" dataset.

While Table III shows a list of top 30 most frequent words for the tweets that include the word "security". This based on the collection of around 7000 tweets that has the word "security" that has been collected over February 24 to April 3, 2016 time period.

TABLE III. THE MOST FREQUENT WORD OCCURS ON THE COLLECTION OF 4100 TWEETS –WITHOUT THE RETWEETS- THAT HAS THE WORD "SECURITY".

Sequence	**Concept**	**Frequency**
1	security	4125
2	privacy	2115
3	twitter	1492
4	hootsuite	1054
5	tweetdeck	580
6	data	560
7	iphone	512
8	amp	442
9	tweet	400
10	tweetbutton	397
11	button	395
12	download	360
13	mobile	333
14	infosec	327
15	android	323
16	facebook	306
17	internet	265
18	cloud	264
19	twitterfeed	260
20	https	254
21	online	253
22	apple	219
23	linkedin	214
24	hipaa	212
25	google	201
26	htc	176
27	apps	174
28	buffer	165
29	bufferapp	165
30	news	163

The last step in showing the project result is to see a different relative topic of privacy and security in their context.

The first explored topic is “HIPAA”. We are going to show the most frequent words occurring when talking about “HIPAA”. Then in the discussion part, we are going to explain the connection between these words and how they are emerging in the context. Using Hootsuit, we have collected around 700 tweets focused on “privacy”, “security”, and “HIPAA” in conjunction. Table IV, Table V, and Table VI present a list of top 30 most frequent words as a result of the frequency results of analyzing the collected tweets after the tweets has been filtered as we did for the above collections. Table IV based on the collection of around 200 tweets that has the word “HIPAASecurity”.

TABLE IV. THE MOST FREQUENT WORD OCCURS ON THE COLLECTION OF 200 TWEETS THAT HAS THE WORD "HIPAASECURITY".

Sequence	Concept	Frequency
1	hipaa	204
2	privacy	187
3	security	182
4	rules	148
5	rule	144
6	hootsuite	230
7	amp	213
8	final	198
9	tweetdeck	193
10	twitter	188
11	web	121
12	breach	120
13	data	110
14	health	103
15	enforcement	96
16	twitterfeed	88
17	ocr	84
18	compliance	77
19	patient	76
20	notification	74
21	hhs	71
22	tweet	66
23	iphone	62
24	button	60
25	tweetbutton	60
26	hitech	58
27	jdsupra	58
28	idexperts	53
29	download	52
30	healthcare	47

Table V based on the collection of around 7000 tweets that has the word “information security”. Table VI based on the collection of around 2000 tweets that has the word "iCloud privacy”. The tables’ information has been collected over February 25 to April 13, 2016 time period.

TABLE V. THE MOST FREQUENT WORD OCCURS ON THE COLLECTION OF 5000 TWEETS –WITHOUT THE RE-TWEETS- THAT HAS THE WORD "INFORMATION SECURITY".I

Sequence	Concept	Frequency
1	Information	5236
2	security	5192
3	online	4047
4	twitter	3057
5	hootsuite	2408
6	twitterfeed	2018
7	facebook	2040
8	amp	1470
9	tweetdeck	1250
10	iphone	1106
11	download	1007
12	data	905
13	protect	900
14	social	899
15	apple	791
16	password	686
17	tips	674
18	microsoft	573
19	tweet	569
20	issue	468
21	computer	458
22	safety	365
23	mobile	363
24	cyber	260
25	help	240
26	android	156
27	media	106
28	dlvr	94
29	guide	53
30	button	22

TABLE VI. THE MOST FREQUENT WORD OCCURS ON THE COLLECTION OF 1919 TWEETS THAT HAS THE WORD "ICLOUD PRIVACY".

Sequence	Concept	Frequency
1	iCloud	2061
2	privacy	2024
3	security	1413
4	twitter	1040
5	twitterfeed	1026
6	hootsuite	990
7	apple	840
8	web	780
9	news	635
10	dlvr	590
11	settings	583
12	amp	482
13	tweetdeck	377
14	unlock	275
15	Encryption	168
16	gremln	153
17	key	148
18	mobile	126
19	iphone	114
20	account	90
21	password	89
22	group	79
23	online	66
24	download	65
25	lock	55
26	apple	51
27	tweet	44
28	bitly	40
29	iOS	30
30	questions	28

Figure 4 represent the relative words in cloud based on their frequency of the “information security” dataset.

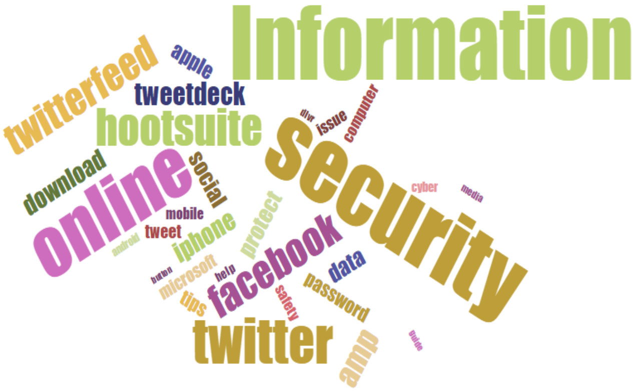

Figure 4. Cloud representation of the relative most frequent words that appears in the "information security" dataset.

We have eliminated lots of unrelated word from the table of the frequency based on the discourse analysis in order to draw a reliable conclusion as seen in Table VII. Table VIII shows different tweets and how they are emerging in the context of privacy.

VI. Discussion

The people view of the privacy and security is reframing due to the rapid growth of social media. The discourse analysis of privacy and security on Twitter revealed the people's understanding of the concepts. The following findings based on the discourse analysis of the collected tweets which further cleared the non-relevant concept of the most frequent word, which served to add more credibility to the collected data. Our analysis discovered that there are similar concepts when Twitter's users are tweeting about the privacy and security. When we used the hash tag to archive the tweets, it wipes out some of the society word that are not a technical from the list of top 30 concepts –in their frequency- which were few, such as surveillance. While including the hash tag make the words more technical related for example, apps. In the context of identity and control each individual's information, the user has the choice to disclose or to keep that information as private. Consequently, they reveal the information based on their trust of the other end who will get the information. When considering the privacy alone without the security, the people more concern about their Facebook and Google personal data privacy. Specially, after the Facebook has changed their term and privacy policy, their users need to be treated respectfully. While in Google the people more concern about Google glass and how it will break the user's privacy which is the most basic right they want to hold.

In addition, Twitter users are care about their mobile data and location privacy by changing their settings. Besides, there is a concern regarding "Consumer Privacy Bill of Rights" in which the users are afraid of losing their rights to maintain their control on the data –which has been collected by any company or organization-. On the other hand, when consider the security with the privacy, we see the same concepts appear in addition to the cloud, HTC, Cyber Intelligence Sharing and Protection Act (CISPA), surveillance, HIPAA, https, information security (infosec), and apps. Thus, in the security context the users care about the cloud and HTC mobile device security and their information security when they were talking about CISPA, surveillance, HIPAA, https -which is uncovering security problem when using Firefox-, and apps. Furthermore, they have talked about how android apps more secure than apple apps. Likewise, they were highlighting that the free apps are associated with significant security risks. In other words, privacy and security are related to government acts and roles, the people freedom, and civil rights [37]. Looking into the HIPAA concepts, we can see clearly how is the user discussing the new added rules and how it will improve the privacy and security of patient records. On the other hand, in the context of Facebook, the users were complaining about Facebook security after "plugs timeline privacy hole" and how they could lock down their account to maximize the account's privacy and security. While in the online context, they were focused more on passwords and giving tips to secure their passwords and make it stronger, and how they can protect their privacy in this era of social media [22]. These results revealed a number of important implications for privacy and security from the tweet's users overview, including the concern regarding the mobile security, Facebook security hole and privacy of their information, privacy of the mobile data and who could access it, the change in regulation and acts which may result in reducing their privacy in this era, privacy and security of their location.

In addition, we found that Twitter is an influential medium that can reflect and educate people regarding the newest news and technology that emerge over the cascading influences of re-tweets and what is the new technology problems, privacy, and security risk that exist with them. Without a doubt, privacy has many elements and factors that formulate the human nature and react to understand fundamental of making the privacy decision [38]. In other words, each person views the privacy and security differently even based on the political representativeness [39].

VII. Recommendations

Looking at the result which includes the twitter users' views, opinions, problems and discussions, we can assist public and private sectors based on our outcome. With the new generation, there are lots of changes in people life. Our data is no longer on hard drive. We moved to cloud base storage which contained a set of sophisticated infrastructures to store data and insure its security and connectivity. In addition, mobile networks have determined ubiquitous connectivity because of the enlarger of users. Beside that comes the smartphones which open the doors to the outsider world through the internet. From the twitter users', we understand that Apple raise their security by adding more sophisticated methods to insure the users' privacy and secure their data, but this advancement has complications in the user end.

TABLE VII. THE MOST FREQUENT WORDS IN ALL OF THE TWEETS COLLECTIONS WHERE THE BOLD BLACK REFERS TO THE SEARCH CRITERIA, THE BLUE COLOR REPRESENT THE UNRELATED CONCEPT TO THE SEARCH CRITERIA, THE GREEN COLOR REPRESENTS HOW OTHER SEARCH CRITERIA RELATED TO THE CURRENT REPRESENTED EACH CRITERIA COLUMN, THE RED COLOR SHOWS THAT THE WORD JUST APPEAR IN ONE COLUMN.

Concept	Privacy	Privacy and security	Security	security and HIPAA	iCloud and Privacy	Information and security
1	**privacy**	**privacy**	**Security**	**hipaa**	**iCloud**	**Information**
2	Robot	**security**	Privacy	**security**	**privacy**	**security**
3	encryption	twitter	twitter	**privacy**	**security**	online
4	download	twitterfeed	hootsuite	rules	twitter	twitter
5	twitter	amp	tweetdeck	rule	twitterfeed	hootsuite
6	Award-winning	hootsuite	Data	hootsuite	hootsuite	twitterfeed
7	hacker	data	iphone	amp	apple	facebook
8	iphone	iphone	Amp	final	web	amp
9	android	download	Tweet	tweetdeck	news	tweetdeck
10	webcams	tweetdeck	tweetbutton	twitter	dlvr	iphone
11	edatesex	android	Button	web	settings	download
12	apple®	dlvr	download	breach	amp	data
13	google	tweet	mobile	data	tweetdeck	protect
14	data	mobile	infosec	health	unlock	social
15	need	htc	android	enforcement	Encryption	apple
16	hootsuite	tweetbutton	facebook	Twitterfeed	gremln	password
17	protector	button	internet	Ocr	key	tips
18	confidential	facebook	Cloud	compliance	mobile	microsoft
19	policy	google	twitterfeed	patient	iphone	tweet
20	**security**	internet	https	notification	account	issue
21	purchase	news	Online	hhs	password	computer
22	strict	https	Apple	tweet	group	safety
23	internet	surveillance	linkedin	iphone	online	mobile
24	mobile	apple	**Hipaa**	button	download	cyber
25	glass	**hipaa**	google	tweetbutton	lock	help
26	online	infosec	Htc	hitech	apple	android
27	products	cloud	Apps	jdsupra	tweet	media
28	tweetdeck	protection	Buffer	idexperts	bitly	dlvr
29	YouTube	cispa	bufferapp	download	iOS	guide
30	app	online	News	healthcare	questions	button

TABLE VIII. SAMPLES OF TWEETS THAT CONTAINING DIFFERENT CONCEPTS FROM THE PRIVACY COLLECTED TWEETS.

Concept	Example
privacy	“Ways to Protect Your Privacy Online - https://t.co/XyQN83caOJ”
Robot	“'Mr. Robot' will take on encryption and privacy in upcoming season https://t.co/sOgSAS4XpJ, see more https://t.co/VKjVZGg7Xx”
encryption	“iPhone encryption: Obama warns against tough stand on data privacy - https://t.co/NAOf3Etmqx ”
twitter	“Twitter receives your information to a bankruptcy, merger, acquisition, reorganization or turn off of Previous Privacy Policy,,FallenSnowden”
hacker	“'Mr. Robot' will take on encryption and privacy in upcoming season - Mashable #tech #technology #hacker #privacy”
iphone	“#FBI could force us to turn on #iPhone cameras and microphones, says #Apple https://t.co/mJidoNAKJ2 #Spying #GovernmentSurveillance #Privacy,,AnthonyLMarin”
android	“How to protect your privacy using Android https://t.co/iXn3u6YSCk https://t.co/TGl7vGGFUk,,drewrobertsmith”
webcams	“Free Cam #Shows on Real #Sex #Webcams”
edatesex	“Free Cam #Shows on Real #Sex #Webcams. Absolute privacy guaranteed - https://t.co/EeOlSHjjMv https://t.co/LSpBKtn3e3,,edatesex”
apple®	“World: We want security and privacy! Blackberry: Here you go! Apple: No, you want this! World: OOH SHINY Blackberry: https://t.co/g7DAehz5gJ,,classam”
Google	“FCC Proposes #privacy Rules for #internet Service Providers https://t.co/PDm5DfLaIF,,vishne0,709202019601416192,443557785,en,"<a href=""http://google.com” “@asymco What price privacy? Apple = your wallet Google = your identity Microsoft = your soul,”
Data	“When talking about data collection issues and privacy, we often forget about those who would willingly donate their data. @DocForeman #SPSM,,chrsmxwll”
Need	“When social media has no longer be your rights and privacy Because need to take care of other's feelings moreeeeeee,,Jifiafa”
hootsuite	“The security vs privacy debate about encryption. https://t.co/BHDsW2LKyL,,renejaspe,709220037693939713,452679265,en,"<a href=""http://www.hootsuite.com”
protector	“Thanks for the free Privacy Screen Protector for my Apple iPhone 6” “Cell Phones : Anti-Spy Peeping Privacy Tempered Glass Screen Protector For iPhone 6 6s 6 Plus …”
confidential	“#Bitcoin privacy: "confidential transactions" feature can fix some of coinjoin's pr...”
Policy	“We have a strict privacy policy”
security	“The Tor Network: The Eyes Hidden Within #technology #security #privacy https://t.co/evmMUoQQwQ #privacy #security #technology #tor,,thinkogram”
purchase	“Kindly note that the purchase is strictly confidential.”
Strict	“You May Soon Be Able to #Restrict How Your #ISP Uses Your #PersonalData https://t.co/ovTSpZ6MuG #privacy #surveillance #DataCollection,,Hfuhs”
internet	“I took back my online privacy with @SurfEasyInc. Encrypt your Internet connection for free with bonus data: https://t.co/nWSjSyXWG8,,TheAzidDawn”
mobile	“How To Preserve Your Privacy When Making Mobile Payments #mobilewebsites https://t.co/M17WnHiImg https://t.co/e7hHnknyf0,,mobilereadywebs”
Glass	“Premium Privacy #Tempered #Glass Screen #Protector for Samsung Galaxy S5 https://t.co/H4QHe9RYrs https://t.co/qLVA5Hhedr,,Markzer1051”
Online	“FOCUS | It's Your Data: Empowering Consumers to Protect Online Privacy https://t.co/cTcujaZZ9Z #ThinkThisThrough.,,ClarkNewhall”
Products	“@smith_nbct I so want to try these! Currently we have privacy concerns with social media but I hope we get there soon #WATeachLead,4687132094,larkscience,709200827441778688,2728553678,en,"<a href=""https://about.twitter.com/products/tweetdeck”
Tweetdeck	“Real #privacy issues not considered in #academic &amp; research worlds | Univ of CA secretly undertook data monitoring https://t.co/4my0dzHOlZ,,frankbaitman,709212521195020288,19314238,en,"<a href=""https://about.twitter.com/products/tweetdeck”
YouTube	“can't wait til that goes live on YouTube what a great explainer about privacy, encryption &amp; cluelessness of the govt &amp; what Apple goes thru”

We need more security as a user, but we do not want our account to be locked while we are working on the phone. Instead of locking Apple ID, they could use the user's finger print as a more convenience way for the users. There is a big part related to the privacy which needs enhancements on the law and regulations. Furthermore, the users need to be educated about their right through video tutorials to guide them over privacy settings instead of long text which usually they do not read it. In short, inform the user of the action and take their choice. For example, Google glass and how the user can protect their identity and information to be easy obtained from such technology. The users have to understand that they should not spread their information on the social media even if it's not complete because the new technologies can link all the data together then have the whole picture. The collected tweets along with their analysis could assist the developer for public and private companies to enhance the privacy and security aspects of their products. Moreover, the researchers can find issue to work on to publish and solve existing problems and open questions.

VIII. Conclusion

With the exponential growth of technology and social media, the privacy has become one of top concern in the digital communication of this era. As the social media technologies developed, the privacy concept expanded. Thus, the privacy is not just about the information control and access. The privacy concept mainly can be viewed in the context of social networks at different levels and categories which are the personal, cultural, and political levels. Privacy from an individual perspective is about how much of the information they can keep without disclosing it. On the other hand, the individual respective of the security come to intersect with the privacy. So, they do not want others to reach their information by using different tools. Even with the security, the people have a concern over their personal information to be reached since there are a security hole or breaches. Consequently, to conceptualizing the privacy from the individual perspective, we have headed to the social media to have a good amount of peoples' thoughts. Thus, we can draw a clearer picture on how the people view the privacy and security concepts. In this project, we have studied and examined the privacy discourses and practices on Twitter, which is one of the leading social media all over the world especially in the USA. In this study, Twitter was the main source of our information. Twitter has a large number of tweets which is posted daily by the Twitter users. The collected texts have served as a database for our discourse analysis on privacy and security for this study. In this project, we have explored the meaning of privacy and security in terms of related dimensions from the Twitter's user's vision. This project significance is to shed a light to see what are the top concerns of the Twitter's users have talked about them. The problem of privacy can be viewed as the relationship of the private realm to the public realm. And as our discourse analysis revealed the privacy of the Twitter's user prospective can be viewed as mixture of cultural and political behaviors and pragmatisms. The people more concern about their Facebook and Google personal data privacy which reveals that people are concerned about their personal information on the social media, which Google and Facebook come into the picture. Beside that the users are anxious about losing their rights to preserve their control of the data which is held by any company or organization. When the security comes to the stage, the gears shift a little to talk about existing laws, technologies, and devices, such as the cloud, HTC, Cyber Intelligence Sharing and Protection Act (CISPA), surveillance, and HIPAA. In this context, the people are going deeper to make decisions when buying a specific device or using explicit technology, for example, when they want to choose a device which use either Android or Apple. When we focus further on Health Information Portability and Accountability Act (HIPAA), we found that the new added rules which will advance the privacy and security of patient records. Beside considerable attention to the privacy breaches and leaks that happen several times. On the other hand, Facebook's users were irritable about the security after "plugs timeline privacy hole" and how they can protect their personal information to be available for unknowing people. In addition to the previous concerns, lots of discussion has been made regarding the security while staying online. Tons of advices have been written in order to assert the people in how to write a strong password. It is particularly significant to note what is the main idea and concern that the users are talking about to consider it in the market world. Our findings suggest deeper implications for analysis the Twitter in term of the concurrency to build up a connected network besides the hosting analysis which will reveal other information more related to our research fields.

References

[1] Jalal, A., & Zaidieh, Y. (2012). The use of social networking in education: Challenges and opportunities. World of Computer Science and Information Technology Journal (WCSIT), 2(1), 18-21. ISSN: 2221-0741

[2] Mourtada, R., Salem, F., Al-Dabbagh, M., & Gargani, G. (2011). The role of social media in arab women's empowerment. Arab Social Media Report, 1(3), 1-26. Retrieved from http://www.arabsocialmediareport.com/UserManagement/PDF/ASMR Report 3.pdf

[3] Da Veiga, A., & Martins, N. (2015). Factorial invariance of an information security culture assessment instrument for multinational organisations with operations across data protection jurisdictions.

[4] Carley, K. M., Columbus, D., & Azoulay, A. (2012). Automap user's guide 2012 (No. CMU-ISR-12-106). carnegie-mellon univ pittsburgh pa inst of software research internat.

[5] Danowski, J. A. (2013). WORDij version 3.0: Semantic network analysis software [computer program]. Chicago: University of Illinois at Chicago.

[6] Salter, M. (2015). Privates in the online public: Sex (ting) and reputation on social media. New Media & Society, 1461444815604133.

[7] Noble, S. U., & Roberts, S. T. (2016). Through google-colored glass (es): design, emotion, class, and wearables as commodity and control.

[8] Solove, D. J. (2015). The meaning and value of privacy. Social Dimensions of Privacy: Interdisciplinary Perspectives, 71.

[9] Al-Saggaf, Y., & Islam, M. Z. (2015). Data mining and privacy of social network sites' users: implications of the data mining problem. Science and engineering ethics, 21(4), 941-966.

[10] Keith, M. J., Thompson, S. C., Hale, J., Lowry, P. B., & Greer, C. (2013). Information disclosure on mobile devices: Re-examining privacy calculus with actual user behavior. International Journal of Human-Computer Studies,71(12), 1163-1173.

[11] Nissenbaum, H. (2010) Privacy in Context. Technology, Policy, and the Integrity of Social Life. Stanford: Stanford Law Books.

[12] Stahl, T. (2016). Indiscriminate mass surveillance and the public sphere.Ethics and Information Technology, 1-7.

[13] Child, J. T., & Starcher, S. C. (2016). Fuzzy Facebook privacy boundaries: Exploring mediated lurking, vague-booking, and Facebook privacy management. Computers in Human Behavior, 54, 483-490.

[14] Kansal, P. (2014). Online privacy concerns and consumer reactions: insights for future strategies. Journal of Indian Business Research, 6(3), 190-212.

[15] Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. Science, 347(6221), 509-514.

[16] Nordgren, A. (2015). Privacy by design in personal health monitoring. Health Care Analysis, 23(2), 148-164.

[17] Charles, V., Tavana, M., & Gherman, T. (2015). The right to be forgotten-is privacy sold out in the big data age?. International Journal of Society Systems Science, 7(4), 283-298.

[18] Lazaro, C., & Le Métayer, D. (2015). Control over Personal Data: True Remedy or Fairy Tale?. SCRIPT-ed, 12(1).

[19] Taddeo, M. (2015). The struggle between liberties and authorities in the information age. Science and engineering ethics, 21(5), 1125-1138.

[20] Xu, H., & Jia, H. (2015). Privacy in a Networked World: New Challenges and Opportunities for Privacy Research. Washington Academy of Sciences. Journal of the Washington Academy of Sciences, 101(3), 73.

[21] Karniel, Y., & Lavie-Dinur, A. (2015). Privacy and Fame: How We Expose Ourselves Across Media Platforms. Lexington Books.

[22] Yao, M. Z. (2011). Self-protection of online privacy: A behavioral approach. In Privacy Online. S. Trepte and L. Reinecke (Eds.). Springer-Verlag Berlin Heidelberg, 111-125.

[23] Rosen, J. (2011). The Unwanted Gaze: The Destruction of Privacy in America. Vintage.

[24] Donath, J., Dragulescu, A., Zinman, A., & Viegas, F. (2010). Data portraits. Leonardo, 43(4), 375-383.

[25] McDonald, P., & Thompson, P. (2016). Social media (tion) and the reshaping of public/private boundaries in employment relations. International Journal of Management Reviews, 18(1), 69-84.

[26] Symmonds, P., Alharbi, A., Nielson, E., & Alshammari, A. (2017). Application of Unsupervised Learning for Detection Cross-site Scripting (XSS) Security Breaches. International Journal of Computer and Information Technology, 6(6), 2279 – 0764.

[27] Klein, B., Laiseca, X., Casado-Mansilla, D., Lopez-de-Ipi˜na, D., & Nespral, A. P. (2012). Detection and extracting of emergency knowledge from twitter streams.

[28] Marcin, I., & Shiu, S. (2012). *Extracting topic trends and connections: Semantic analysis and topic linking in twitter and wikipedia datasets*. Informally published manuscript, Computer Science, Stanford University, Retrieved from http://www.stanford.edu/class/cs224w/upload/cs224w-004-final.v01.pdf

[29] Zubiaga, A., Spina, D., Martinez, R., & Fresno, V. (2015). Real-time classification of twitter trends. *Journal of the Association for Information Science and Technology*, *66*(3), 462-473.

[30] Parra, D., Trattner, C., Gómez, D., Hurtado, M., Wen, X., & Lin, Y. R. (2016). Twitter in academic events: a study of temporal usage, communication, sentimental and topical patterns in 16 computer science conferences. *Computer Communications*, *73*, 301-314.

[31] Bonchi, F., Castillo, C., Gionis, A., & Jaimes, A. (2011). Social network analysis and mining for business applications. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), Article 22. doi: 10.1145/1961189.1961194

[32] Okazaki, S., Díaz-Martín, A. M., Rozano, M., & Menéndez-Benito, H. D. (2015). Using Twitter to engage with customers: a data mining approach.*Internet Research*, *25*(3), 416-434.

[33] Java, A., Song, X., Finin, T., & Tseng, B. (2009). Why we twitter: An analysis of a microblogging community .*Advances in Web Mining and Web Usage Analysis In Advances in Web Mining and Web Usage Analysis, Vol. 5439 (2009), pp. 118-138, doi:10.1007/978-3-642-00528-2_7 Key: citeulike:4510863*, *5439*, 118-138. doi: 10.1007/978-3-642-00528-2_7

[34] Alsoos, M., & Kheirbek, A. (2015). A Semantic Approach to Enhance HITS Algorithm for Extracting Associated Concepts using ConceptNet. *Journal of Digital Information Management*, *13*(1), 53.

[35] Chhabra, S., Aggarwaly, A., Benevenutoz, F., & Kumaraguru, P. (2011). Phi.sh/$ocial: The phishing landscape through short urls. *ACM*, 92-101. doi: 978-1-4503-0788-8

[36] Russell, M. G., Flora, J., Strohmaier, M., oschko, J. P., Perez, R., & Rubens, N. (2011). Semantic analysis of energy-related conversations in social media: A twitter case study. *ACM*.

[37] Min, J. (2016). Personal information concerns and provision in social network sites: Interplay between secure preservation and true presentation. *Journal of the Association for Information Science and Technology*, *67*(1), 26-42.

[38] Yuan, E. J., Feng, M., & Danowski, J. A. (2013). "Privacy" in semantic networks on Chinese social media: The case of Sina Weibo. *Journal of Communication*, *63*(6), 1011-1031.

[39] Horne, C., Darras, B., Bean, E., Srivastava, A., & Frickel, S. (2015). Privacy, technology, and norms: The case of Smart Meters. *Social science research*, *51*, 64-7