Privacy and Security on Twitter: A Semantic Network Analysis

Running head: Privacy and Security on Twitter

INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 1)
Available online at www.ijariit.com

Privacy and Security on Twitter: A Semantic Network Analysis

Mansour A. Alshaikhsaleh
mansouralshaikh@oakland.edu
Oakland University, Rochester, Michigan, USA

Dr. Faisal Abdullah
abdullfa@lewisu.edu
Lewis University, Romeoville, Illinois, USA

Prof. Mohamed A. Zohdy
zohdyma@oakland.edu
Oakland University, Rochester, Michigan, USA

Ahmad Alshammari
alshammari@oakland.edu
Oakland University, Rochester, Michigan, USA

Exclusive Summary

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 view 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 project and draw our conclusions. We have collected more than 125000 tweets to analysis them based on different aspects.

This project mainly focused on privacy and security then it takes into account the Health Information Portability and Accountability Act (HIPAA), Facebook, and online 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 Facebook and Google personal data privacy. Particularly, when 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. In addition, the Twitter's users were emphasizing that the free apps are linked with substantial security risks.

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, Facebook's users were uncomfortable and complaining about the security after "plugs timeline privacy hole". They 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.

These 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.

Table of Contents

Introduction 6
Statement of the Problem 6
Definition of Terms 6
Methodology 7
1. Automated Part 8
2. Manual Part 8
Research Question 9
Significance of the Study and Potential outcomes 9
Literature Review 10
Theories of Privacy 10
Empirical Studies of Analyzing Twitter 15
Results 19
Discussion 31
Conclusion 34
References 37

Table of Figures

Figure 1: Relative percentage cross texts for the most frequent word occurs on the collection of 100000 tweets that has the word "privacy" 19

Figure 2: Cloud representation of the relative most frequent words that appears in the "privacy" dataset 20

Figure 3: Cloud representation of the relative most frequent words that appears in the "privacy" and "security" dataset 23

Figure 4: Cloud representation of the relative most frequent words that appears in the "privacy", "security", and "Facebook" dataset 27

Table of Tables

Table 1: The most frequent word occurs on the collection of 100000 tweets that has the word "privacy" 20

Table 2: The most frequent word occurs on the collection of 20915 tweets that has the word "privacy" and "security"..... 21

Table 3: The most frequent word occurs on the collection of 3792 tweets that has the word "privacy" and "security" with hashtag 23

Table 4: The most frequent word occurs on the collection of 727 tweets that has the word "privacy", "security", and "HIPAA" 24

Table 5: The most frequent word occurs on the collection of 1002 tweets that has the word "privacy", "security", and "facebook"..... 25

Table 6: The most frequent word occurs on the collection of 1114 tweets that has the word "privacy", "security", and "online" 26

Table 7: The most frequent words in all of the tweets collections where the bold black refers to the search criteria, the white and gray 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..... 28

Table 8: Samples of tweets that containing different concepts from the privacy collected tweets 29

Introduction

Statement of the Problem

As the new generation moving toward technology, the social media has played an important role in reflecting the people ideas and thoughts without any bias (Jalal & Zaidieh, 2012). 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 (Jalal & Zaidieh, 2012). We can extract valuable information from different social media by analyzing their data (Mourtada, Salem, Al-Dabbagh & Gargani, 2011); 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."

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 (Margulis, 2003). 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 (Pfleeger & Pfleeger, 2007).

Methodology

i. Data collection which has been done by searching the tweets which contain privacy and security key word using TwapperKeeper which now called HootSuite Archives:
    1. The first set includes all the tweets that have both keywords "privacy".
    2. The second set includes all the tweets that have both keywords "privacy" and "security".
    3. The third set includes more specific tweets which have both keywords "#privacy" and "#security".
    4. The fourth set includes all the tweets that have the following keywords "privacy", "security", and "HIPAA".
    5. The first set includes all the tweets that have that have the following keywords "privacy", "security", and "Facebook".
    6. The first set includes all the tweets that have that have the following keywords "privacy", "security", and "Online".

ii. Data analysis where we have used AutoMap to create a concept list which a list of each word and its frequency of occurrence in the collected tweets.

iii. Data visualization as a final step which we have used **many eays for IBM** to visualize our finding and how the Twitter users' relate different keywords to the security and privacy terms.

We have two major parts in our study which are:

1. Automated Part

In particular, using semantic network analysis software such as AutoMap (Carley, Columbus, & Azoulay, 2011) and WORDij (Romppel, 2011; Yuan, 2013), 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 called Manyeyes. 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. Nissenbaum, 2004) or "surveillance" (i.e. Gandy, 1993; Solove, 2007).

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 strengthen the findings obtained in the quantitative steps of this research project.

Research Question

Q1 - what are the semantic networks of "privacy" and "security" that develop on Twitter?

Q2 - What keywords and key-phrases intersect with "privacy" and "security" on Twitter?

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.

Literature Review

Theories of Privacy

In modern Western societies, privacy is often recognized as a basic condition for individual autonomy, liberty, identity, and integrity (Altman, 1975; Nissenbaum, 2010; Rachels, 1975; Solove, 2001; Westin, 1967). The concept of privacy has been thoroughly explored in a number of 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. Westin (1967) intended privacy as a dynamic process of access control, and a fundamental means to personal autonomy, emotional release, self-evaluation, and limited and protected communication (Margiulis, 2012). Altman (1975) 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 Altman's dialectic understanding of privacy, Petronio (2002) explained the process of boundaries negotiation through the communication management theory (CMT). CMT is based on the assumption 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 (Petronio, 2002). In addition, privacy emerges is a bi-directional process of access negotiation that operates at the individual and at the group level (Altman, 1975; Margiulis, 2012). Thus, privacy may be intended as an individual value, a societal value (Regan, 1995), and a cultural universal (Altman, 1975).

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. Margulis, 2003; Rachels, 1975). 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. Traditional theories of privacy have also adopted such a distinction (Fried, 1990; Moor, 1997). A common approach suggests that privacy has instrumental value, as it is necessary to protect one from specific practical harms (Fried, 1990; Dwyer, 2009; Waldo et al., 2007). To claim that privacy has higher instrumental values, one should prove the role of privacy in achieving higher-level benefits. Rachels (1975), for example, argued that privacy is necessary to maintain a variety of relationships.

Suggesting the intrinsic value of privacy, Johnson (1994) explained that privacy is a fundamental aspect of autonomy. Similarly, Moor (1997) 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 a "natural expression of the need of security" (p. 29). Privacy, in other words, protects us from strangers. Floridi (2005) further supported the intrinsic value of privacy suggesting an ontological interpretation of it. He argued that 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 (Floridi, 2005).

Theories of privacy have approached it from normative perspectives (Warren & Brandeis, 1890; Westin, 1967) or adopting merely descriptive approaches (Gavison, 1980). Overall, discussions around privacy have entailed a variety of dimensions that include psychological,

legal, social, and informational elements (Yao, 2011). To address privacy, Nissenbaum (2010) 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 (2010) 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 that of freedom (Post, 2001; Rosen, 2011). 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 (as in Donath et al., 2010). Privacy, instead, enables one to manage the flow of information and avoid misrepresentation and distorted knowledge (Rosen, 2011; Solove, 2007). 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" (Donath et al., 2010, p. 375). Problems emerge when such a portrait differ from desired representations of the self (Solove, 2007). 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 (Post, 2001). 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 (Fried, 1990; Post, 2001; Rosen, 2011). 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" (Post, 2001, p. 2092) and constitute an intrinsic offense against one's dignity. Such an offense is harmful as it hinders one's ability to save public face (Whitman, 2004; Solove, 2007). Dignity, in this context, relates to the respect that individuals owe to each other as members of a community and depends on the respect of shared social norms (Post, 2001). "Privacy as dignity safeguards the socialized aspects of the self" (p. 2095) emphasizing the importance of one's reputation to develop identity and to partake in different communities. 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 (Whitman, 2004).

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" (Rosen, 2011, p. 11). This sense of privacy stems from a concept of freedom that originated in the Renaissance or in the Reformation (Berlin, 1958). The understanding of privacy as freedom entails the negotiation between social norms and

self-interest, but focuses on the latter. "Privacy as freedom safeguards the spontaneous, independent, and uniquely individual aspects of the self" (Post, 2001, p. 2095) placing the emphasis on the individual over the community. Adopting this approach, some have intended privacy as freedom from external social norms epitomized in the family (Sennet, 1977). Others have interpreted it as a liberty from state regulation and from intrusion within the domestic walls (Rosen, 2011; Whitman, 2004). The latter is often associated with the approach adopted in American laws that points to the "value of liberty against the state within the privacy of one's home" (Whitman, 2004, p. 1214).

Analyzing the value of liberty, Berlin (1958) 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. Berlin (1958) noted that 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 (Margiulis, 2003).

Pos's (2001) classification of the three concepts of privacy will be the main theoretical lens used to inform the current study.

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 (Klein, Laiseca, Casado-Mansilla, Lopez-de-Ipi˜na & Nespral, 2012).

Marcin and Shiu (2012) 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 a 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.

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 (Bonchi, Castillo, Gionis & Jaimes, 2011). Jansen, Zhang, Sobel & Chowdury (2009) have analyzed over 150,000 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. After word, they were able to identify that 19% of the tweets have mentioned the brand name. 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 (Java, Song, Finin & Tseng, 2009). They have used a two-level approach combining HITS algorithm (Kleinberg, 1999) 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 (Java, Song, Finin & Tseng, 2009).

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 (Chhabra, Aggarwaly, Benevenutoz & Kumaraguru, 2011).

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 (Russell, Flora, Strohmaier, oschko, Perez & Rubens, 2011).

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 thought 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 (Bonchi, Castillo, Gionis & Jaimes, 2011).

Results

The first step in conducting the project result is the quantitative analysis. Using Hootsuit, we have collected more than 100000 tweets focused on “privacy”. Figure 1 presents a bar graph of relative percentage for each word in the top 60 most frequent cross texts. This based on the collection of 100000 tweets that has the word "privacy" that has been collected over February 24 to March 14, 2013, ten days as the time period.

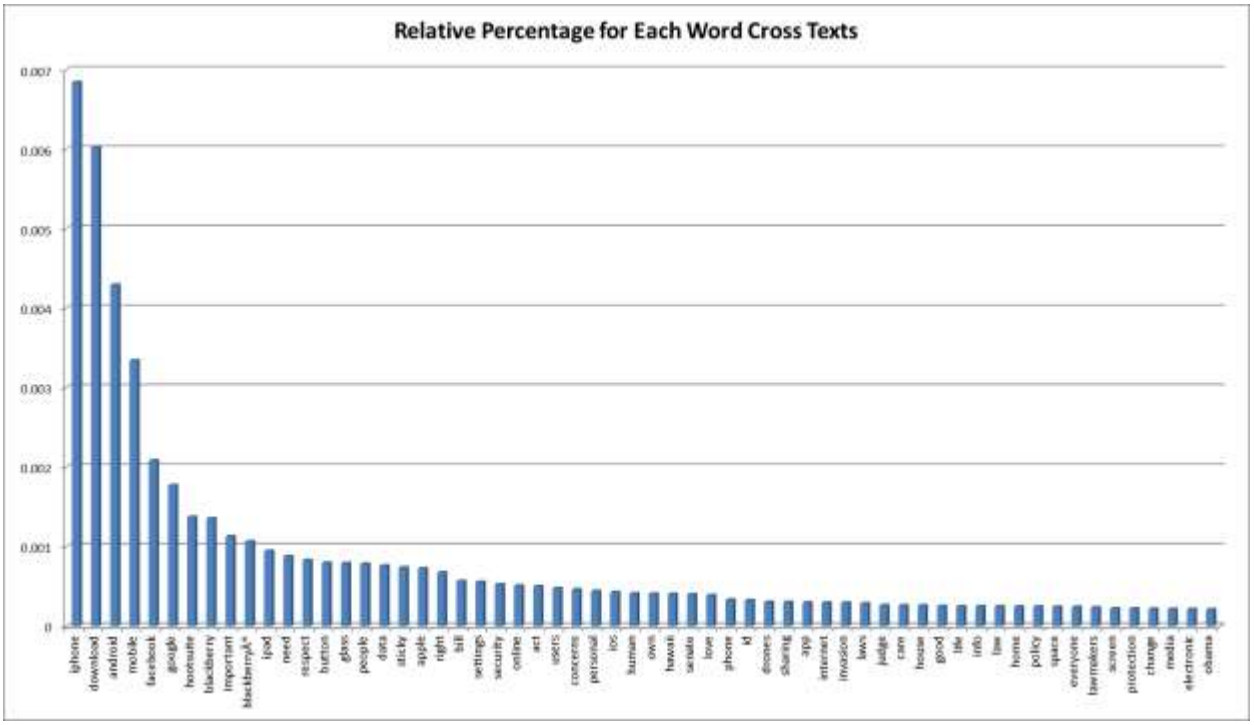

Figure 1: Relative percentage cross texts for the most frequent word occurs on the collection of 100000 tweets that has the word "privacy"

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

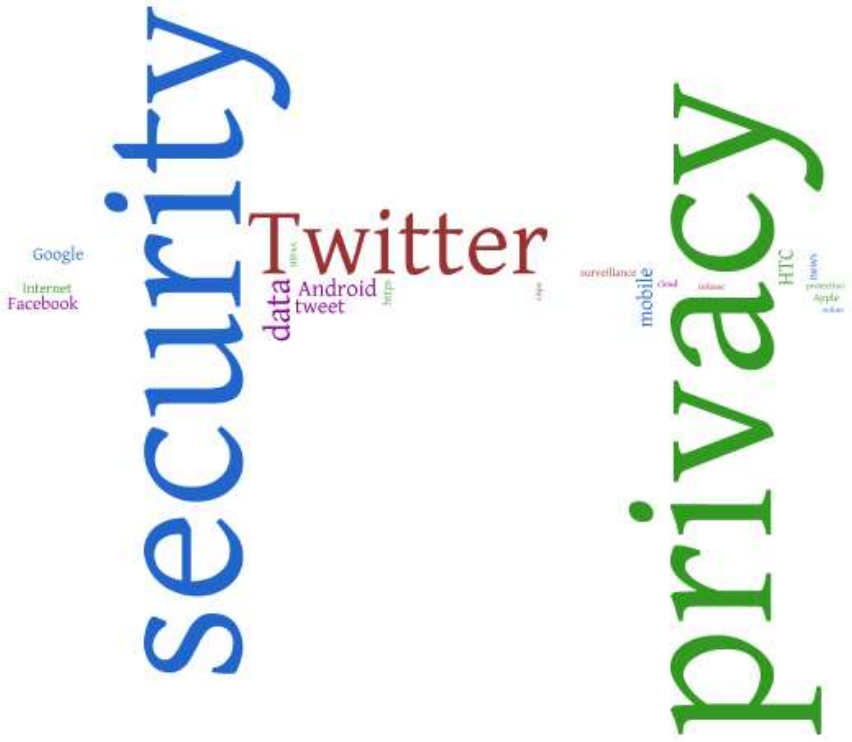

Figure 2: Cloud representation of the relative most frequent words that appears in the "privacy" dataset

In the following Table 1 which 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 1300 words.

Table 1: The most frequent word occurs on the collection of 100000 tweets that has the word "privacy"

Sequence	Concept	Frequency
1	twitter	82143
2	iphone	32617
3	download	28715
4	android	20471
5	mobile	15930
6	facebook	9928
7	google	8448
8	hootsuite	6543

9	blackberry	6461
10	important	5381
11	blackberryÂ®	5070
12	ipad	4503
13	need	4183
14	respect	3952
15	button	3791
16	glass	3774
17	people	3729
18	data	3629
19	sticky	3519
20	apple	3458
21	right	3213
22	bill	2697
23	settings	2651
24	security	2504
25	online	2426
26	act	2376
27	users	2268
28	concerns	2195
29	personal	2100
30	ios	2031

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 20000 tweets focused on “privacy” and “security” in conjunction.

Table 2: The most frequent word occurs on the collection of 20915 tweets that has the word "privacy" and "security"

Sequence	Concept	Frequency
1	privacy	18638
2	security	17805
3	twitter	8508
4	twitterfeed	5651
5	amp	3626

6	hootsuite	3497
7	data	3132
8	iphone	2875
9	download	2586
10	tweetdeck	2277
11	android	2051
12	dlvr	1920
13	tweet	1894
14	mobile	1856
15	htc	1851
16	tweetbutton	1778
17	button	1776
18	facebook	1596
19	google	1572
20	internet	1260
21	news	1117
22	https	1034
23	surveillance	972
24	apple	929
25	hipaa	817
26	infosec	800
27	cloud	790
28	protection	757
29	cispa	728
30	online	715

Table 2 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 3 shows a list of top 30 most frequent words for the tweets that include the word “privacy” and “security” with the hostages. This based on the collection of around 3700 tweets that has the word "privacy" and “security” that has been collected over February 24 to April 3, 2013 time period.

Table 3: The most frequent word occurs on the collection of 3792 tweets that has the word "privacy" and "security" with hashtag

Sequence	Concept	Frequency
1	privacy	4125
2	security	4115
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 4, 5, and 6 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 4 based on the collection of around 7000 tweets that has the word "privacy", “security”, and “HIPAA”.

Table 4: The most frequent word occurs on the collection of 727 tweets that has the word "privacy", "security", and "HIPAA"

Sequence	Concept	Frequency
1	hipaa	804
2	privacy	658
3	security	592
4	rules	248
5	rule	244
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 5 based on the collection of around 1000 tweets that has the word "privacy", “security”, and “facebook”. Table 6 based on the collection of around 1100 tweets that has the word "privacy", “security”, and “online”. The tables’ information has been collected over February 25 to April 3, 2013 time period.

Table 5: The most frequent word occurs on the collection of 1002 tweets that has the word "privacy", "security", and "facebook"

Sequence	Concept	Frequency
1	facebook	1236
2	privacy	992
3	security	747
4	twitter	357

5	twitterfeed	248
6	hootsuite	208
7	timeline	204
8	web	147
9	news	125
10	dlvr	116
11	settings	107
12	amp	105
13	tweetdeck	100
14	saranonline	99
15	plugs	91
16	gremln	86
17	fixes	74
18	mobile	73
19	iphone	69
20	account	68
21	hole	68
22	group	65
23	online	63
24	download	60
25	hacker	60
26	apple	56
27	tweet	56
28	bitly	54
29	socialmedia	53
30	fraud	52

Table 6: The most frequent word occurs on the collection of 1114 tweets that has the word "privacy", "security", and "online"

Sequence	Concept	Frequency
1	privacy	761
2	security	724
3	online	713
4	twitter	340
5	hootsuite	326
6	twitterfeed	190
7	facebook	144
8	amp	138
9	tweetdeck	135
10	iphone	90

11	download	83
12	data	82
13	protect	77
14	social	75
15	apple	68
16	password	68
17	tips	68
18	microsoft	66
19	tweet	64
20	issue	60
21	computer	59
22	safety	59
23	mobile	56
24	cyber	55
25	help	55
26	android	51
27	media	51
28	dlvr	50
29	guide	50
30	button	48

Figure 4 represent the relative words in cloud based on their frequency of the "privacy", “security”, and “Facebook” dataset.

Figure 4: Cloud representation of the relative most frequent words that appears in the "privacy", "security", and "Facebook" 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 7. Table 8 shows different tweets and how they are emerging in the context of privacy.

Table 7: The most frequent words in all of the tweets collections where the bold black refers to the search criteria, the white and gray 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	Privacy and security with hashtag	Privacy, security, and HIPAA	Privacy, security, and facebook	Privacy, security, and online
1	**privacy**	**privacy**	**privacy**	**hipaa**	**facebook**	**privacy**
2	twitter	**security**	**security**	**privacy**	**privacy**	**security**
3	iphone	twitter	twitter	**security**	**security**	**online**
4	download	twitterfeed	hootsuite	rules	twitter	twitter
5	android	amp	tweetdeck	rule	twitterfeed	hootsuite
6	mobile	hootsuite	data	hootsuite	hootsuite	twitterfeed
7	facebook	data	iphone	amp	timeline	facebook
8	google	iphone	amp	final	web	amp
9	hootsuite	download	tweet	tweetdeck	news	tweetdeck
10	blackberry	tweetdeck	tweetbutton	twitter	dlvr	iphone
11	important	android	button	web	settings	download
12	blackberryÂ®	dlvr	download	breach	amp	data
13	ipad	tweet	mobile	data	tweetdeck	protect
14	need	mobile	infosec	health	saranonline	social
15	respect	htc	android	enforcement	plugs	apple
16	button	tweetbutton	facebook	Twitterfeed	gremln	password
17	glass	button	internet	Ocr	fixes	tips
18	people	facebook	cloud	compliance	mobile	microsoft
19	data	google	twitterfeed	patient	iphone	tweet
20	sticky	internet	https	notification	account	issue
21	apple	news	online	hhs	hole	computer
22	right	https	apple	tweet	group	safety
23	bill	surveillance	linkedin	iphone	online	mobile
24	settings	apple	hipaa	button	download	cyber
25	security	hipaa	google	tweetbutton	hacker	help
26	online	infosec	htc	hitech	apple	android
27	act	cloud	apps	jdsupra	tweet	media
28	users	protection	buffer	idexperts	bitly	dlvr
29	concerns	cispa	bufferapp	download	socialmedia	guide
30	personal	online	news	healthcare	fraud	button

Table 8: Samples of tweets that containing different concepts from the privacy collected tweets

Concept	Example
privacy	“Facebook loses privacy rights case & must set up the Digital Trust Foundation as part of settlement”
twitter	“How #Twitter Knows So Much About You Before You've Even Signed Up” “Twitter: 80% of Government User Data Requests Issued Without a Warrant #twitter #data #privacy #security”
iphone	“#tech #apps #iphone Privacy bill would ban police from getting email and location data without a warrant” “Don't Be Dumb About #Smartphone #Privacy http://t.co/ENCWPBpsn0 #mobile #iphone” “If you prefer an individual's level of privacy tremendously, then you should probably use Siri less. http://t.co/4twX9a75gY # #iphone”
download	“#download New Approaches for Security, Privacy and Trust in Complex Environments http://t.co/dZgznWcQ0I”
android	“iOS apps leak more personal data than do Android apps http://t.co/wXl2zWxpHT #app #ios #android #security #leaks #data #privacy” “79% of the top 50 free #iOS and #Android apps are associated with risky behaviors or #privacy issues.” “Access files on locked, encrypted #Android phones by putting them in a freezer for an hour” http://t.co/lLM1rFj2Sb #hacki #privacy”
mobile	“In Pictures: 10 Incredibly Simple Things You Can Do To Protect Your Privacy ... Privacy & data management on #mobile devices … ” “Guidance on How #Mobile #Apps Can Better Protect Consumer #Privacy”
facebook	“you should never marry your #twitter acct. to your #Facebook acct. you'll never get any social privacy...” “Facebook has broadly decreased users' privacy, against their preferences, 7-yr study”
google	“I bet if Google or Facebook were run by idealistic college kids they would respect privacy more.”
hootsuite	rel=""""nofollow""""">twitterfeed"",http://a0.twimg.com/profile_images/803188566/ExpertAutoIns_normal.jpeg,,0,0," Mar 10 21:50:25 +0000 2013"",1362952225"
blackberry	“Blackberry Removes Four Apps from Store for Privacy... http://t.co/bblIDKDBRO #IndustryNews #Blackberry #InMobile #DigitalnaSigurnost”
important	“Guy can’t walk in Seattle bar wearing Google Glasses: Privacy is important to many maybe this will become a trend!”
ipad	If you prefer an individual's level of privacy tremendously, then you should probably use Siri less. http://t.co/4twX9a75gY #i #iphone
need	“My privacy about the only thing I need back”
respect	“Stop dont go there this is my private square s-t-o-p please respect my privacy” “respect my privacy please !!!”
button	“ … >Tweet Button< …”
glass	“Google Glass: is it a threat to our privacy?: The tech giant's 'wearable computing' project is now being teste...” “Google glass technology prompts major privacy concerns”
people	“More people are opting to #travel by #PrivateJet for convenience, comfort, safety and privacy.”
data	“Frozen #Android phones give up data secrets: http://t.co/Jftj3McSGM Frost - Forensic Recovery of Scrambled Telephones. #Security #Privacy”
sticky	“ … http://a0.twimg.com/sticky/default_profile_images/default_profile_1_normal.png ...”
apple	“Apple can't duck judge's iPhone privacy suit orders - Pittsburgh Post Gazette: Apple can't duck judge... http://t.co/bY3luTkF #privacy”
right	“There should be policy to protect the right to privacy with this new technology”
bill	“A bill called CISPA may erase every online privacy law ever written.” “New bill to protect your webmail and location #privacy needs your support: The Electronic Communications Privacy...”
settings	“Facebook went public, because even they couldn't figure out the Privacy Settings.”

	"I feel more real using twitter cuz I always use privacy settings for almost every single post in Facebook. 2 people can see 2 different me."
security	"How a friend's hacked Facebook Account can compromise your privacy and security"
online	"Can States Do A Better Job Of Protecting Online #Privacy?"
act	"it's because of some privacy act they have for the kids but we have an early meeting at the school"
users	" … Facebook Users are Sharing More, Despite Network Privacy Concerns: Regardless how much people complain about"
concerns	"Contentious development in ed tech for K - 12, will certainly raise privacy concerns. Worth a read!"
personal	"Despite Efforts To Protect Privacy, Users Increase Amount of Personal Information Available Online"

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 (Westin, 2003).

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 (Yao, 2011).

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. 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 and security decision (Solove, 2002). In other words, each person view the privacy and security differently even based on the political representativeness (Kasper, 2005).

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

Altman, I. (1975). *The environment and social behavior*. Selmont, CA: Wadsworth Publishing Company.

Berlin, I. (1958). Two Concepts of Liberty.

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

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

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

Farmer, D., & Mann, C. C. (2003). Surveillance nation. *Technology Review*, *106*(3), 34–40.

Floridi, L. (2005). The ontological interpretation of informational privacy. *Ethics and Information Technology*, 7(4), 185-200.

Foucault, M. (1995). *Discipline and Punish: The Birth of the Prison*. New York: Vintage Book.

Fried, C. (1990). Privacy: A rational context. In M.D. Ermann, M.B. William, & C. Gutierrez (Eds.), *Computers, ethics, & society*. New York: Oxford University Press, Inc.

Gandy, O.H. (1993). The Panoptic Sort: A Political Economy of Personal Information. Boulder, CO: Westview Press

Gavison, R. (1980). Privacy and the limits of the law. *Yale Law Journal, 89*, 421-471.

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

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009).Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology* 60(11):2169–2188.

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

Kasper, D. V. S. (2005). The evolution (or devolution) of privacy. *Sociological Forum*, *20*, 69-92.

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.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, *46*(5), 604–632.

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

Margulis, S. T. (2003). Three theories of privacy: An overview. In S. Trepte & L. Reinecke (Eds.), *Privacy Online.* Retrieved from https://www.filepicker.io/api/file/7gAK71IIQneaKdZeHbVq

Moor, J. (1997). Towards a theory of privacy in the information age. *ACM SIGCAS Computers and Society, 27*(3), 27-32.

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

Nissenbaum, H. (1998). Protecting privacy in an information age: The problem of privacy in public. *Law and Philosophy 17*(5), 559–596.

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review, 79*, 119-158.

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

Petronio, S. (2002). *Boundries of privacy: Dialectics of disclosure*. Albany, NY: State University of New York Press.

Pfleeger, C. P., & Pfleeger, S. L. (2007). *Security in computing*. (Fourth ed., p. 880). Prentice-Hall PTR. Retrieved from http://my.safaribooksonline.com/book/networking/security/0132390779

Post, R. C. (2001). Three concepts of privacy. *Faculty Scholarship Series*. Paper 185. Retrieved from http://digitalcommons.law.yale.edu/fss_papers/185

Rachels, J. (1975). Why privacy is important. (1975). *Philosophy and Public Affairs*, *4*(4), 323-333.

Regan, P. M. (1995). *Legislating privacy, technology, social values, and public policy*. Univ of North Carolina Pr.

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

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*.

Solove, D. J. (2002). Conceptualizing privacy. *California Law Review*, 90 (4), 1088–1154.

Solove, D. J. (2007). *The Future of Reputation: Gossip, Rumor, and Privacy on the Internet*. Yale University Press.

Warren, S. D. & Brandeis, L. D. (1890). The right to privacy. *Harvard Law Review 4*(5), 193–220.

Westin, A. F. (1967). *Privacy and Freedom*. Atheneum, New York.

Westin, A. F. (2003). Social and political dimensions of privacy. *Journal of Social Issues, 59*, 431-453.

Whitman, J. Q. (2004). The two western cultures of privacy: Dignity versus liberty. *Yale Law Journal*, 1151-1221.

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.