ASIA-AFRICA JOURNAL OF RECENT SCIENTIFIC RESEARCH (AAJRSR) VOL. 2, 2022
A PUBLICATION OF IAPAAR IN COLLABORATION WITH MEDI-CAPS UNIVERSITY, INDIA
ISSN: 2814-0397 Available online at https://journals.iapaar.com/index.php/AAJRSR

DATA MINING: THE THREAT TO THE PRIVACY OF FACEBOOK SOCIAL NETWORK SITE USERS

Md Zahidul Islam
Centre for Research in Complex Systems,
School of Computing and Mathematics,
Charles Sturt University, Bathurst, NSW 2795, Australia

Abstract

This article explores the potential of data mining as a technique that could be used by malicious data miners to threaten the privacy of Facebook Social Network Site users. It applies a data mining algorithm, specifically a clustering algorithm, to a hypothetical dataset of a sample of individuals from Saudi Arabia, Pakistan and Yemen to show the ease at which characteristics about the SNSs users can be discovered and used in a way that could invade their privacy. It is hoped by exploring the threats from data mining on individuals' privacy is enable SNSs developers better understand the ways in which SNSs can be used by malevolent data miners to harm users and how to operate in SNS safely. Another important outcome is to help developers of SNSs develop mechanisms that will provide protection to users from these knowledge discovery techniques.

Keywords: data mining, privacy, facebook, social network

Introduction

The web environment has changed dramatically in the last few years with the emergence of one of the most fertile environments of personal data on the web; social network sites (SNSs). Social Network Sites (SNS) continue to be among the most popular websites on the internet. According to most recent rankings from Alexa (2011) of the top 500 sites, Face- book is ranked second from the top (in terms of the total number of page views) followed by YouTube in third place, Blogger in the seventh and Twitter in the tenth; suggesting that social networking is one of the popular internetactivities among the two billion world internet citizens (Internet World Stats 2011). There are many SNS on the web including MySpace, Twitter, Hi5, Flickr, Orkut, LinkedIn and BeBo, but by far Facebook is the most popular site (Alexa 2011). Indeed, this observation is consistent with Facebook statistics that showed in less than a year, the number of active users on Face- book has jumped from 500 million to 800 mil- lion users, 50 per cent of whom log on to it each day (Facebok 2011). This massive growth in the population size of Facbook is indicative of the huge popularity that Facebook enjoys. A social network site (SNS), defined formally, is a web-based service that allows individuals to "1) construct a public or semi-public profile within a bounded system, 2) articulate a list of other users with whom they share a connection, and 3) view and traverse their list of

connections and those made by others within the system (Boyd and Ellison, 2007). Users are also prompted to invite other SNSs users to view and follow their personal profile and web behaviour within the SNSs and to label and categorise these relationships with other users as "friends" (Facebook) or sometimes "followers" (Twitter). This definition is not just chosen because it is the most frequently used one but also because it incorporates most of the elements found on Facebook such as being a platform where individuals can communicate with each other to form new social connections and/or to maintain existing friendships and in doing so share personal information, photos, videos, thoughts as well as their feelings. This view about Facebook is in line with the results from a recent study which has shown that people use SNS not only to develop new friendships but also to communicate with older friends whom they cannot meet regularly face-to-face (Al-Saggaf 2011). In addition, Facebook allows its members to communicate with others using voice, videos, online chat, offline messages, blogs and 'wall-ing'[1] (Al-Saggaf 2011). Upon registering with Facebook, the first thing users do is create their own profiles, which they can set to be either private or public. Like any other SNS, Facebook allows its users to upload their photos, videos and emotional states. The site also allows outside developers to build applications which users can then use to personalise their profiles and perform other tasks, such as comparing movie preferences and charting travel histories (Boyd and Ellison 2007). In addition, users can also create their own online identities (Jones et al 2008). To create their identities on SNS, users need to place on their own profiles their personal and biographical data such as name (a real name or alias), date and place of birth, citizenship, nationality, photos, hobbies, and so on. While on the one hand users are increasingly aware and very concerned about their privacy on Facebook (Al-Saggaf 2011), on the other hand, self-disclosure and revealing private information on the site is very widespread (Al-Saggaf 2011) with users sharing with strangers up-to-the-minute updates of the status of their feelings and thoughts. Given young adults between the ages of 18-24 represent the largest cohort of SNS users (Hoy & Milne 2010), revealing private information such as political views, residential address, date of birth, books read, movies watched, schools went to, sexual orientations and their inner thoughts about their partners, neighbours, colleagues and employers can have serious consequences for these users' informational privacy and possibly also for their financial and physical security.

With hundreds of millions of web users around the world now uploading masses of personal data onto the web within the confines of personal networks of friends, family, and acquaintances, the success of social network sites, such as Facebook and Twitter, has made a wealth of social networking data available for businesses and more malicious data miners to profit from. Major culprits include banks and credit agencies, and advertising companies, who are usually assisted in their data mining practices by SNSs with whom they are affiliated. Sar and Al-Saggaf, (2013) study has found that by just visiting non SNSs and SNSs, third party sites can track the users' activities by the use of the HTTP cookies. These third party sites can be advertising sites or data aggregators, or other SNSs. However, these sites are not only collecting these massive amounts

of data about SNSs users; they are also mining them, for the purpose of placing the users in new and non-obvious classifications or categories that the users themselves did not know they ever existed. The aim of this article is to explore the potential of data mining clustering algorithm as a technique that could be used by malicious data miners to threaten the privacy of Facebook SNSs users.

Privacy and Data Mining

Privacy on Facebook can be threatened in many ways including through the continuous changing of the privacy settings that Facebook does without announcing to users, tracking technologies such as HTTP cookies that gather information about users without their knowledge and Application Programming Interface (API) tools that enable other SNS to share users' information and create complete profiles enabling them, in essence, to sell this information to third parties. Another method that can be used to erode users' privacy is data mining. When a data mining algorithm is applied to a large dataset which can be created by harvesting users' information from SNS, hidden and non-obvious patterns about those users can be unearthed from this dataset (Tavani 2011). Used unethically on Facebook users, data mining can have the potential of incorrectly placing users in newly created categories or groups that could, for example, make them victims of organised crime. But what is privacy? What is considered as private? And why is privacy valued anyway? There are three theories of privacy (Tavani 2011). The first theory relates to the notion of Accessibility Privacy, which defines privacy in terms of one's physically 'being let alone' or freedom from intrusion into one's physical space. The second theory relates to Decisional Privacy, which defines privacy in terms of freedom from interference in one's choices and decisions. The third, and most relevant theory to this article, is that of Informational Privacy, which defines privacy as control over the flow of one's personal information, including the transfer and exchange of that information (Tavani 2011). In addition, Moor (2000) has also introduced an account of privacy that is more comprehensive in that it encapsulates all these three theories. Specifically, it incorporates the elements of the non-intrusion, non-interference and informational views of privacy.

According to Moor, an 'individual has privacy in a situation if in that particular situation the individual is protected from intrusion, interference, and information access by others. Tavani also notes that Moor makes a distinction between naturally private and normatively private situations. According to Tavani (2011), this distinction allows us to distinguish between the conditions required for having privacy (in a descriptive sense) and having a right to privacy in the normative sense. According to this distinction, if a person sees another picking her nose in the library, then that person lost her privacy but her privacy was not violated. But if that other person peeps through the keyhole of her apartment door then her privacy is not only lost but also violated. Privacy is valued for many reasons including for achieving important human

ends like trust and friendships (Tavani 2011). For Jim Moor (2004, cited in Tavani 2011), privacy is important because it is the articulation or expression of the core value of security (ibid). On the other hand, for Deborah Johnson privacy is an important social good because when people are watched all the time they take the perspective of the observer. Because decisions will be made on the basis of what they do, they tend to think before acting. This becomes a form of social control which leads to eroding individuals' freedom. This in turn affects democracy (Johnson 2001). In our view, privacy is also important for love, marriage and partner relationships, respect and dignity, freedom of expression and liberty, autonomy, solitude, anonymity and secrecy, data protection and self-confidence to name a few. Erosion of privacy due to excessive self-disclosure is a problem for participants in SNS and users are more than ever concerned about their privacy (Al-Saggaf 2011). That said, self-disclosure is rampant on both SNS (Park and Kee 2009) and online communities (Rheingold 2000). There are many reasons for this but the lack of oral and non-verbal cues, and lack of public self-awareness are major factors. Lack of oral and non-verbal cues, and lack of public self-awareness cause abandonment of social inhibitions and detachments from social conventions (Barnes 2001).

Trust between online communicators has been found to be another factor that encourages self-disclosure (Valenzuela, Park and Kee 2009). Trust is vital for personal relationships (Weckert 2003); in fact, one way to show the strength of a friendship between two individuals is by demonstrating they trust each other. At the same time to show trust in someone I have to reveal more about myself. That is why self-disclosure is also important for personal relationships (Rheingold, 2000). Indeed, researchers have found that as people become familiar with others online, they tend to reveal more about themselves (Barnes 2001). But, as online communicators reveal more and more sensitive information about themselves, the chances that their privacy will be eroded or, at least, violated will be increased. For example, women in Saudi Arabia who place their photos (even the ones that show their faces) on Facebook may become subject to sexual coercion by malicious SNS users if these photos fall into these malicious SNS users' hands. They could threaten to release their photos or video clips on the internet or publicly via Bluetooth on mobile phones unless these women submit to their demands. If the women do not comply, the result can cause serious damage to their family's reputation which is a grave matter in Saudi society. While this example suggests that privacy is not valued equally in all cultures (placing photos on Facebook may not be an issue for women in Western societies), it nevertheless shows that privacy has at least some value in all societies (Tavani 2011).

Data mining has traditionally been restricted to trolling through offline 'data warehouses', large storehouses of personal data consisting mainly of commercial transactions (Tavani, 2011). However, a profitable application of data mining, especially in recent years, has been in all sorts of web-based services. The World Wide Web provides web data miners with a virtually limitless

repository of information on which to employ their data gathering techniques. Vast amounts of personal information (e.g. names, addresses, web browsing behaviour, personal preferences, online consumer records, etc) are made publically searchable through the content of web pages. Web data mining techniques can be employed to automatically discover, extract, and analyse information from web documents and services to generate new and financially lucrative data (Kosala and Blockeel, 2000). There are three types of web data mining. The first type is web structure mining, which allows miners to extract the links and structure of websites for the purpose of discovering previously unknown relationships between the mined websites. Applying this technique results in a structural summary of the websites or web pages (Ting, 2008). The second type is text mining, known also as web content mining (Kosala and Blockeel, 2000). This technique involves the scanning and extracting of contents from webpage, i.e. text, pictures, graphs, audio, video, and hyperlinks, for the purpose of determining the relevance of the contents to a search query (Kosala and Blockeel, 2000). One of the main uses of this method is gathering and summarising the best possible information available from web content for a user requesting that information (Ting, 2008). The third type is web usage mining (Ting, 2008), which involves tracking and analysing the navigation behaviour of web users for the purpose of understanding the movements of users online and what they are doing (Van Wel and Royakkers, 2004). One of the main uses of this method is to generate data that represents, without identifying the individual, the web users' actions in a public environment (Van Wel and Royakkers, 2004). Web usage mining can identify traces such as an IP-address of a computer, type of browser used, date and time of login to a site, hyperlinks followed, use of cookies with frequent visits to particular sites, etc, can be extracted using web usage mining techniques to discover patterns in web browsing behaviour. Web usage mining has the potential to be privacy violating due to the fact that all internet service providers (ISPs) can link a web users' navigation behaviour to their personal information, which the user has had to provide to their ISP in order for the ISP to supply the user access to the internet (Van Wel and Royakkers, 2004).

Applications of Data mining

Data mining has a wide range of applications such as social network analysis, analysis of microarray gene expression data, soft- ware engineering, market segmentation, web search result grouping, and irrigation water demand forecasting – just to name a few (Zhao and Zhang 2011). Data pre-processing techniques can be applied to collect data from various sources including weather stations, satellite images and water usage statements. Various other data mining techniques (such as classification by decision tree and artificial neural network) can then be applied on the collected data to extract patterns that can then be used to predict future water demand for irrigation. Accurate water demand prediction can help us to manage irrigation water more efficiently by reducing the wastage of already limited irrigation water. Data mining can also be applied on huge amounts of data generated from social networking

sites (SNS) in order to study contactnetworks, growth rates and social implicationsof various SNS (Patton 2007).

Data mining is also used in direct marketing (where targeted advertisements are sent to potential customers), and various business analyses. An example of its application in developing better business strategies and business pro- motion is association rule mining (Islam 2008). Association rule mining is applied on a huge setof transactions (records), where each transaction consists of a list of items such as milk, bread, butter that are sold together in the transaction. A supermarket chain, for example Wal-Mart, collects each and every sale transaction in a central data warehouse from all its stores (around 2900) situated in around six different countries. From the collected data, they can explore frequent item sets and association rules. When a set of items appear in a huge number of trans-actions then they are known as a frequent item set. Moreover, when the appearance of one set of items in a transaction increases the possibility of the appearance of another set of items in the same transaction then it is known as association rule. Based on frequent item sets and association rules a supermarket can design its store in order to increase sales. For example, they can put the frequent items apart in a store so that customers need to walk through the store to find the items. This can often increasethe sales as customers tend to buy more itemswhen they walk through them. Moreover, supermarkets can reduce the prices of a set of items and advertise the price reduction. However, they can increase the prices of another set of items when they know from association rulemining that the increased sale of the former set of items will also increase the sales of the later sets of items and therefore, they will makemore overall profit.

The common steps in the data collection and analysis process: data preparation; collection of raw data until extraction process (e.g. crawler) concludes its cyclical sampling; cleaning of data to make sure there is no duplication of information; and analysis of data gathered, usually represented in graph form, with 'nodes' of a graph representing users and the structure of the graph representing the relationships between users.

Data mining and threats to the privacy of users of Facebook

Data mining can extract various interesting andhidden patterns (such as logic rules and clusters) from a dataset. Often it is argued that general patterns (logic rules) are public knowledge and should be considered sensitive to individual privacy (Islam and Bronkovic, 2011). For example, the pattern that people from Asiahave black hair is public knowledge and there-fore should not be considered sensitive. If a per- son is originally from Asia then it is generally assumed that he/she has black hair and there is nothing wrong with that. However, some general patterns even when they are applicable to many people can disadvantage an individual and therefore need to be considered as sensitive. For example, if a bank discovers that all previous loans granted to the customers livingin a specific suburb are defaulted and, there- fore, decides to turn down the loan application of an individual living in the suburb

just based on his/her address then this pattern can appear to be unfair to the individual. Similarly, if a business learns from its past records that all employees of a race have been found inefficient and, therefore, decides not to offer a job to a new applicant who is from the same race, then the pattern can again appear to be unfair to the applicant. Tavani (2011) presents an example of such intrusive and privacy threatening data mining where a car loan application, for purchasing a new BMW, was turned down by a bank although the applicant was financially solvent and had an executive position with good salary. The bank mines its data sets and discovers a pattern that executives earning between $120K and $150K annually, who often take expensive vacations and purchase luxury cars generally start their own businesses. The bank then mines another data set and finds that the majority of such entrepreneurs declare bankruptcy within a year. Unfortunately, the applicant for the car loan falls in this category, although neither the applicant nor the bank knows whether the applicant will really start his own business and face bankruptcy in future.

Data mining can be used for extracting various sensitive patterns of activity by the Facebook or other SNS users. The patterns can then be applied on any Facebook (or any other SNS) user resulting in huge user discomfort and serious breach of individual privacy. For example, a malicious data miner can study the Face- book activities of his/her friends. Based on the observation he/she can then prepare a dataset (a table with records and attributes) having information on the users. In the dataset each record can represent a Facebook user and each column/attribute can represent a property of the user. The properties can be learnt from the Facebook use patterns. Examples of the attributes can include: 'Ratio of the number of opposite sex friends to same sex friends' (Col. 6 in Table 1), 'Number of own picture uploads per week' (Col. 3 in Table 1), 'Level of exposure of the pictures (i.e. how exposed the person is in the pictures)' (Col. 4 in Table 1), and 'Number of status changes in Facebook per week' (Col. 5 in Table 1). There are various data mining tasks including data collection (such as automatic collection of weather data from different sensors), data pre-processing (such as pre-processing of satellite images), data integration/transformation to prepare a suitable flat file containing a range of relevant non-class attributes (such as blood sugar level and cholesterol level of a patient) and a class attribute (such as diagnosis of disease), data cleansing (by identifying and removing/correcting corrupt data, and imputing missing values), and pattern extraction from the clean pre-processed data (Islam 2012). There are a number of pattern extraction techniques including clustering (such as K-means and Fuzzy C-means clustering).

ISSN: 2814-0397 Available online at https://journals.iapaar.com/index.php/AAJRSR

Table 1: Sample dataset of Facebook users

Age	Nationality	Own Pic/week	Exposure	Status/ week	Ratio
21	Saudi	7	High	3	0.1
19	Pakistan	2	High	12	5
33	Yemen	7	High	2	4
5	Saudi	2	High	7	0.1
19	Pakistan	2	Medium	12	3
25	Yemen	3	Medium	3	4
22	Saudi	7	Medium	7	0.1
33	Pakistan	8	Medium	8	5
21	Saudi	7	Medium	22	5
27	Saudi	8	Medium	33	1
22	Saudi	2	Low	22	7
26	Saudi	8	Low	2	2
21	Saudi	7	High	3	0.1
19	Pakistan	2	High	12	5
33	Yemen	7	High	2	4
35	Saudi	2	High	7	0.1
19	Pakistan	2	Medium	12	3
25	Yemen	3	Medium	3	4
22	Saudi	7	Medium	7	0.1
33	Pakistan	8	Medium	8	5
21	Saudi	7	Medium	22	5
27	Saudi	8	Medium	33	1
22	Saudi	2	Low	22	7
26	Saudi	8	Low	2	2

Hypothetical dataset is presented in Table 1, created from Facebook activities and supplementary knowledge of a malicious data miner. A data miner can apply an existing clustering algorithm on the dataset in order to cluster the records (users) where similar records are grouped together in one cluster and dissimilar records are grouped together in different clusters. In order to simulate the actions of a malicious data miner we implement an existing Fuzzy c-means clustering technique (Lee and Pedrycz, 2009), which has been shown in the literature to be better than many other existing techniques, and then apply the technique on our example dataset and thereby produce three clusters.

Table 2: Results of the application of clustering on the sample dataset of Facebook users

Cluster	Age	Nationality	Own Pic/ week	Exposure	Status/ week	Ratio
Cluster 1	19	Pakistan	2	High	12	5
	19	Pakistan	2	Medium	12	3
	25	Yemen	3	Medium	3	4
	22	Saudi	2	Low	22	7
	19	Pakistan	2	High	12	5
	19	Pakistan	2	Medium	12	3
	25	Yemen	3	Medium	3	4
	22	Saudi	2	Low	22	7
Cluster 2	22	Saudi	7	Medium	7	0.1
	33	Pakistan	8	Medium	8	5
	21	Saudi	7	Medium	22	5
	27	Saudi	8	Medium	33	1
	26	Saudi	8	Low	2	2
	22	Saudi	7	Medium	7	0.1
	33	Pakistan	8	Medium	8	5
	21	Saudi	7	Medium	22	5
	27	Saudi	8	Medium	33	1
	26	Saudi	8	Low	2	2
Cluster 3	21	Saudi	7	High	3	0.1
	33	Yemen	7	High	2	4
	35	Saudi	2	High	7	0.1
	21	Saudi	7	High	3	0.1
	33	Yemen	7	High	2	4
	35	Saudi	2	High	7	0.1

Table 2 present the result of the application of clustering on the sample dataset of Facebook users. The data miner can then study the properties of the records belonging to a cluster. For

example, a common property of Cluster 1 is having high number of opposite sex friends compared to the number of same sex friends. Some other common properties are having reasonably high number of Status updates per week, and low number of own picture uploading per week.

However, some common properties of Cluster 3 are quite opposite where the users have generally low ratio of opposite sex friends to same sex friends, low number of status updates per week, but high number of own picture uploading per week where the uploaded pictures have high exposure.

Such an observation may make the data miner more curious about the friends falling in Cluster 3 resulting in more careful analysis and follow up observations on the activities (both online and offline) of the friends. Therefore, falling in a group with such patterns can cause an uncomfortable situation for a female in a conservative society like the Saudi society where such a breach of private information of a female member of a family can cause serious damage to the reputation of the family.

Moreover, since the dataset has been created based on the Facebook information of the friends (female) of the data miner, he/she is likely to have supplementary knowledge on the users. By a careful analysis (based on his/her supplementary knowledge) of the friends falling in Cluster 3 the data miner may come to a conclusion that these friends are generally less conservative and "Willing to date". At this stage the data miner may reasonably come to a conclusion that any female Facebook user having similar properties is possibly also "Willing to date". In data mining, this is generally known as labelling the records with class values. Based on this understanding, any unknown Facebook users who have similar properties as the properties of Cluster 3 can be categorised as "Willing to date" (in data mining, this is often called classification) and therefore approached inappropriately by the malicious data miner resulting in huge social discomfort for the user.

Conclusion and Suggestions

This paper revealed the potential of data mining as a technique that could be used by malicious data miners to threaten the privacy of Facebook SNS users. Using a hypothetical dataset of a sample of individuals from Saudi Arabia, Pakistan and Yemen, a data mining clustering algorithm was applied. The application of the clustering algorithm on the above hypothetical dataset has shown that the threats from data mining on individuals' privacy can be serious. A malicious data miner can cluster his/her friends into different groups. The miner can then analyze the properties of the friends belonging to a group. Additionally, from his supplementary knowledge on the friends he can also add a label on them. Based on the extracted properties of the friends and the label he/she can next classify other Facebook users (who are not even in his/her friend list) into the label resulting in a breach of privacy. In this study we show an

example where the data miner can classify other users into "Willing to Date" due the similarity of their properties and the properties of a group of his friends. In this example, the label "Willing to Date" is his supplementary knowledge on a group of his friends. Note that this is just an example of a possible attack from a malicious data miner on the individual privacy of the Facebook users. There are many other possible ways to attack the privacy of the Facebook users. In this study we suggest that, to preserve privacy online, it is important that users mask their data or hide some information such as date of birth, address and other identifying information. This way, even if an unknown malicious data miner can classify a user, for example, as 'Willing to date', it can be difficult for the miner to locate the user and thus harass her. At the same time, if the properties of a cluster are known to the users they can deliberately design their activities in a way so that she is not classified as 'Willing to date'. For example, if a user (originally belonging to Cluster 3) has Exposure = "High" then she can deliberately change it to "Low" in order to avoid being grouped with other users in Cluster 3. In sum, if all (or most) SNSs users place some false and misleading information about themselves (as we are suggesting), then a data miner will face difficulties to conclude with certainty whether or not the SNSs users really have a connection with a particular cluster.

Privacy is often not taken seriously by many users especially the young users group. It is essential for SNS users to be careful in maintaining their privacy online for reasons discussed above. Therefore, it is important to raise their awareness about the possible privacy invasions and their implications. One way to preserve privacy online is by masking the individual's data carefully or hiding the sensitive information such as date of birth, address and other identifying information. This way, even if an unknown malicious data miner can classify a user, for example, as 'Willing to date', it can be difficult for the miner to locate the user and thus harass her. At the same time, if the logic rules are known to the users they can deliberately design their activities in a way so that she is not classified as 'Willing to date'. Another way to protect privacy online is by put- ting restrictions, possibly in the form of laws and regulations, on the use of data mining for individual purposes. However, restricting malicious data mining by introducing laws can be a difficult job. First it will be difficult to detect that someone is mining data maliciously. Second, even if detection is possible it will be difficult to prove malicious intent. One can always argue that he/she was performing data mining for good intentions such as research and knowledge discovery.

Lastly it is suggested that the data mining community should develop Privacy Preserving Data Mining (PPDM) techniques specifically catered for online environments. Until such techniques are developed, users should protect themselves by using all possible ways including hiding their identifying information, using SNS privacy settings carefully and masking their online activities to protect them from being identified as potential victims. Ensuring the privacy settings on Facebook are up to date (often they roll back to the default settings) can be another way to keep malicious data miners out of the way.

References

Alexa (2011a). The top 500 sites on the web. Available online at http://www.alexa.com/topsites,

Alexa (2011b) The top 500 sites on the web. Available online at http://www.alexa.com/topsites/countries/SA,

Al-Saggaf, Y. (2003) Online communities in Saudi Arabia: An ethno- graphic study, PhD thesis, Charles Sturt University, Wagga Wagga, Australia

Al-Saggaf, Y. (2006) The online public sphere in the Arab world: The war in Iraq on the Al Arabiya website, Journal of Computer-Medi- ated Communication, Vol. 12, No. 1. Available online at http://jcmc. indiana.edu/vol12/issue1/al-saggaf.html

Al-Saggaf, Y. (2011) Saudi females on Facebook: An ethnographic study, International Journal of Emerging Technologies and Society, Vol. 9, No. 1 pp 1-19

Al-Saggaf, Y. and Weckert, J. (2011) Privacy from a Saudi Arabian perspective, Journal of Information Ethics, Vol. 20, No. 1 pp 34-53

Barnes, S. B. (2001). Online connections: Internet interpersonal rela- tionships, Hampton Press, New Jersey

Bastani, S. (2000) Muslim women on-line, Arab World Geographer, Vol. 3, No 1 pp 40-59

Boyd, D. M. (2006) Friends, friendsters, and Top 8: Writing com- munity into being on social network sites, First Monday, Vol. 11, No. 12. Available online at http://www.firstmonday.org/issues/ issue11_12/boyd

Boyd, D. M. and Ellison, N. B. (2007) Social network sites: Definition, history, and scholarship, Journal of Computer-Mediated Communi- cation, Vol. 13, No. 1. Available online at http://jcmc.indiana.edu/ vol13/issue1/boyd.ellison.html

Boyd, D. M. and Ellison. N. B. (2007) Social network sites: Definition, history, and scholarship, Journal of Computer-Mediated Communication, 13: 210-230.

Burmeister, O. K., Weckert, J. and Williamson, K. (2011) Seniors extend understanding of what constitutes universal values, Journal of Information, Communication and Ethics in Society, Vol. 9, No. 4. pp 238-252

Cocking, D. and Matthews, S. (2000) Unreal friends, Ethics and Information Technology, Vol. 2 pp 223-231

Dyson, E. (1998) Release 2.1: A design for living in the digital age, Broadway Books, New York

Emerson, D. (2008) Facebook friends not real friends: judge, Sydney Morning Herald. Available online at http://www. smh.com.au/news/technology/facebook-friends-not-real-judge/2008/03/27/1206207279597.html

Facebook (2011) Statistics. Available online at http://www.face-book.com/press/info.php?statistics

Fule, P. and Roddick, J. F. (2004) Detecting privacy and ethical sensi- tivity in data mining results. Paper presented to the 27th Australian Computer Science Conference (ACSC2004), Vol. 26 of Conference in Research and Practice in Information Technology, Estivill-Castro, Vladimir (ed.) pp 163-168

Garton, L., Haythornthwaite, C. and Wellman, B. (1997) Studying online social networks, Journal of Computer-Mediated Communi- cation, Vol. 3, No. 1. Available online at http://jcmc.indiana.edu/ vol3/issue1/garton.html

Hamman, R. (2001) Computer networks linking network communi- ties, Werry, C. and Mowbray, M. (eds) Online communities: Com- merce, community action, and the virtual university, Hewlett-Pack- ard, New Jersey pp 71-95

Hauben, M. and Hauben, R. (1997) Netizens: On the history and impact of usenet and the internet, IEE Computer Society Press, Washington

Haythornthwaite, C. and Wellman, B. (2002) The internet in every- day life: An introduction, The Internet in Everyday Life, Wellman, B. and Haythornthwaite, C. (eds) Blackwell Publishers, Oxford pp 1-55

Horn, S. (1998) Cyberville: Clicks, culture, and the creation of an online town, Warner Books, New York

Hoy, H.G and Milne, G. 2010. Gender differences in privacy-related measures for young adult Facebook users, Journal of Interactive Advertising, Vol. 10 pp 1525-2019

Huang, D., and Pan, W. (2006) Incorporating biological knowledge into distance-based clustering analysis of microarray gene expres- sion data, Bioinformatics, Vol. 22 pp 1259-1268

Internet World Stats (2011) Internet usage in the Middle East. Available online at http://www.Internetworldstats.com/stats.htm,

Islam, M. Z. (2008) Privacy preservation in data mining through noise addition. PhD thesis in Computer Science, School of Electri- cal Engineering and Computer Science, the University of Newcastle, Australia

Islam, M. Z. (2012) Explore: A Novel Decision Tree Classification Algorithm, Lecture Notes in Computer Science, Vol. 6121 pp 55-71

Islam, M. Z. and Brankovic, L. (2011) Privacy preserving data min- ing: A noise addition framework using a novel clustering technique, Knowledge-Based Systems, December, Vol. 24, No. 8 pp 1214-1223

Johnson, D. G. (2001) Computer ethics, New Jersey, Prentice Hall, third edition

Joinson, A. (1998) Causes and implications of disinhibited behaviour on the internet, Psychology and the Internet, Gackenbach, J. (ed.), Academic Press, San Diego pp 43-60

Jones, S. G. (1998) Information, internet, and community: Notes toward an understanding of community in the information age, CyberSociety 2.0: Revisiting Computer-Mediated Communication and Community, Jones, S. G. (ed.), Sage Publications, Thousand Oaks, CA pp 1-35

Jones, S., Millermaier, S., Goya-Marthinez, M. and Schuler, J. (2008) Whose space is MySpace? A content analysis of MySpace profiles, First Monday, Vol. 13, No. 9. Available online at http:// firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view-Article/2202/2024

Karpinski, A. C. and Duberstein, A. (2009) A description of Facebook use and academic performance among undergraduate and gradu- ate students, Technology Research Poster Session. Available online at http://researchnews.osu.edu/archive/facebook2009.jpg

Kollock, P. and Smith, M. (1999) Communities in cyberspace, Com- munities in cyberspace, Smith, M. and Kollock, P. (eds) Routledge, London pp 3-25

Kosala, R., and Blockeel , H. (2000). Web mining research: A survey, SIGKDD Explorations, 2: 1-15.

Lange, P. G. (2007) Publicly private and privately public: Social net- working on YouTube, Journal of Computer-Mediated Communi- cation, Vol. 13, No. 1. Available online at http://jcmc.indiana.edu/ vol13/issue1/lange.html

Lee, M. and Pedrycz, W. (2009) The fuzzy c-means algorithm with fuzzy P-mode prototypes for clustering objects having mixed features. Journal of Fuzzy Sets and Systems, 160 (24): 3590-3600

Lung, C.-H., Zaman, M., and Nandi, A. (2004) Applications of cluster- ing techniques to software partitioning, recovery and restructuring, Journal of Systems and Software, Vol. 73, pp 227-244

Mar, J. (2000) Online on time: The language of internet relay chat, Gibbs, D. and Krause, K. L. (eds) Cyberlines: Languages and cultures of the internet, James Nicholas Publishers, Australia pp 151-174

Markham, A. N. (1998) Life online: Researching real experience in virtual space, AltaMira Publications, Walnut Creek, CA

Mitra, A. (1997) Virtual commonality: Looking for India on the internet, Jones, S. G.(ed.) Virtual culture: Identity and communica- tion in cybersociety, Sage Publications, London pp 55-79

Moor, J. (2000) Towards a theory of privacy for the information age, Baird, R. M., Ramsower, R. and Rosenbaum, S. E. (eds) Cybe- rethics: Moral, social, and legal issues in the computer age, Pro- metheus Books, New York pp 2000-2012

Moor, J. (2004) Reason, relativity, and responsibility in computer ethics, Spinello, R. and Tavani, H. T. (eds) Readings in cyberethics, Jones and Bartlett Publishers, Sudbury, MA, second edition pp 40-54

Muralidhar, K., Parsa, R., and Sarathy, R. (1999) A general additive data perturbation method for database security, Management Sci- ence, Vol. 45, No. 10 pp 1399-1415

Patton, S. (2007) Social Networking Sites: Data Mining and Investi- gative Techniques. Available online at https://www.blackhat.com/ presentations/bh-usa-07/Patton/Whitepaper/bh-usa-07-patton-WP. pdf

Preece, J. (2000) Online communities: Designing useability, support- ing sociability, John Wiley and Sons, Chichester

Rafaeli, S. and Sudweeks, F. (1997) Networked interactivity, Jour- nal of Computer-Mediated Communication, Vol. 2, No. 4. Available online at http://jcmc.indiana.edu/vol2/issue4/rafaeli.sudweeks.html,

Rheingold, H. (2000) The virtual community: Homesteading on the electronic frontier, MIT Press, Cambridge, revised edition

Rifkin, J. (2000) The age of access: How the shift from ownership to access is transforming capitalism, Penguin Books, London

Sar, R. K. and Al-Saggaf, Y. (2013). Social networking sites' tracking of unintentionally shared information. First Monday, 18(6).

Tamura, T. (2005) Japanese feeling for privacy. Proceedings of the 2nd Asia Pacific Computing and Philosophy Conference, Honglada- rom, S. (ed.) Novotel Hotel, Bangkok, Thailand, January pp 88-93

Tavani, H. T. (2011) Ethics and technology: controversies, questions, and strategies for ethical computing, Hoboken, N. J., John Wiley, third edition

Tavani, H.T. (1999) Informational privacy, data mining, and the Internet, Ethics and Information Technology, 1: 137-145.

Ting, I (2008). Web mining techniques for on-line social networks analysis, International Conference on Service Systems and Service Management, June 30 2008-July 2: 1-5.

Tsai, C. Y., and Chiu, C. C. (2004) A purchase-based market segmen- tation methodology, Expert Systems with Applications, Vol. 27 pp 265-276

Valenzuela, S., Park, N. and Kee, K. F. (2009) Is there social capital in a social network site? Facebook use and college students' life sat- isfaction, trust, and participation, Journal of Computer-Mediated Communication, Vol. 14 pp 875-901

Van Wel, L. and Royakkers, L.(2004) Ethical issues in web data mining, Ethics and Information Technology, 6: 129-140.

Wallace, P. (1999) The psychology of the internet, Cambridge Uni- versity Press, Cambridge

Weckert, J. (2003) On-line trust, The impact of the internet on our moral lives, Cavalier, R. (ed.) Suny Press, Albany, NY pp 95-117

Weckert, J. and Adeney, D. (1997) Computer and information eth- ics, Westport, Connecticut/London, Greenwood Press

Wellman, B. and Gulia, M. (1999) Net-surfers don't ride alone: Vir- tual communities as communities, Wellman, B. (ed.) Networks in the global village: Life in contemporary communities, Westview, Colorado pp 331-366

Young, Y. (2009) Online social networking: An Australian perspec- tive, International Journal of Emerging Technologies and Society, Vol. 7, No. 1 pp 39-57

Zamir, O., and Etzioni, O. (1999) Grouper: a dynamic clustering interface to Web search results, Computer Networks: The Interna- tional Journal of Computer and Telecommunications Networking, Vol. 31 pp 1361 – 1374

Zhao, P. and Zhang, C. Q. (2011) A new clustering method and its application in social networks, Pattern Recognition Letters, Vol. 32 pp 2109 - 2118