This article is downloaded from

http://researchoutput.csu.edu.au

It is the paper published as:

Author: M. Z. Islam and Y. Al-Saggaf  
Title: Privacy in Social Network Sites (SNS) - the threats from Data Mining  
Journal: Ethical Space: The International Journal of Communication Ethics  
ISSN: 1742-0105  
Year: 2012  
Volume: 9  
Issue: 4  
Pages: 32 - 40  
Abstract: This paper explores the potential of data mining as a technique that could be used by malicious data miners to threaten the privacy of SNS users and makes a moral case for the users’ right to privacy. It applies a data mining 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 SNS 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 and arguing for users' right to privacy, the study will raise SNS users' awareness about the ways in which information that they reveal online can be used by malevolent data miners to harm them and how to operate in SNS safely.


Author Address: zislam@csu.edu.au  
yalsaggaf@csu.edu.au

CRO Number: 41035
Privacy in Social Network Sites (SNS): the threats from data mining

Yeslam Al-Saggaf and Md Zahidul Islam

School of Computing and Mathematics, Charles Sturt University, 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 SNS users and makes a moral case for the users’ right to privacy. It applies a data mining algorithm to a hypothetical dataset developed for the purpose of this article, to show the ease at which characteristics about the SNS 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 and arguing for users’ right to privacy, the study will raise SNS users’ awareness about the ways in which information that they reveal online can be used by malevolent data miners to harm them and to how to operate in SNS safely.

Introduction: SNS and Facebook

Social Network Sites (SNS) continued to be among the most popular websites on the internet. According to most recent rankings from Alexa (2011a) of the top 500 sites on the internet, Facebook is ranked second from the top (in terms of the total number of page views) followed by Youtube in the third, Blogger in the seventh and Twitter in the tenth position; suggesting that social networking is one of the popular internet activities among the two billion (2,095,006,005) 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 2011a). Indeed this observation is consistent with Facebook statistics that showed in less than a year, the number of active users on Facebook has jumped from 500 million to 800 million users, 50% of which log on to it each day (Facbok 2011). This massive growth in the population size of Facebook is indicative of the huge popularity that Facebook enjoys.

There are not many definitions of SNS in the literature (Al-Saggaf 2011), but the most frequently cited definition is the one proposed by Boyd & Ellison (2007: 211) who defined SNS as Web-based
services that allow 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. This definition is not just chosen because it is the most frequently 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 ‘walling’ (Al-Saggaf 2011).

Upon registering in 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 & 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 any other information members would like to post on their profiles.

While on one hand users are increasingly aware and very concerned about their privacy on Facebook (Boyd & Ellison 2007, Jones et al 2008, Young 2009, Al-Saggaf 2011), on the other hand, self disclosure and revealing private information on the site is very widespread (Jones et al 2008, Valenzuela, Park & Kee 2009, Al-Saggaf 2011) with users sharing with strangers up to the minute updates of the status of their feelings and thoughts. Given the majority of Facebook users are from the

---

1 The participants in Al-Saggaf (2011) study call the activity of posting comments and other objects on their Facebook wall to communicate with their friends as ‘walling’.
youth as several studies have shown (see, for example, Jones et al 2008, Lenhart 2009, Valenzuela, Park & Kee 2009, Young 2009), 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.

This article explores the potential of data mining as a technique that could be used by malicious data miners to threaten the privacy of SNS users and makes a moral case for the protection of users’ privacy. By exploring threats from data mining on individuals’ privacy and arguing for users’ right to their privacy, the study will raise SNS users’ awareness about the ways in which information that they reveal online can be used by malevolent data miners to harm them. To achieve this aim, a hypothetical dataset will be created. Next, the article will apply a data mining algorithm to this hypothetical dataset to demonstrate the ease at which characteristics about the SNS users can be discovered and used in a way that could invade their privacy. After that the article will present a short philosophical analysis to argue for the importance of protecting users’ privacy from the threats of data mining. At the end, the article will present several recommendations that should contribute to raising users’ awareness about how to operate in SNS safely.

**Privacy as an Ethical Issue**

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 like 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 about these users enabling them in essence to sell this information to third parties. Another method that can be used to erode user’s 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 (Bronkovic, Islam & Giggins
Privacy is one of the most widely discussed topics in the Australian media and in the computer and information technology ethics literature (Al-Saggaf and Weckert 2011). But what is privacy? What is considered as private? And why is privacy valued anyway? There are three theories of privacy (Tavani 2012). The first theory is the 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 is the 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 the Informational Privacy, which defines privacy as control over the flow of one's personal information, including the transfer and exchange of that information (Tavani 2012, p. 137). In addition, Moor’s (cited in Tavani 2012) has also introduced an account of privacy that is more comprehensive in that it encapsulated all these three theories. Specifically, it incorporated the elements of the non-intrusion, non-interference, and informational views of privacy (Tavani 2012).

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 2012, 137). Tavani also notes that Moor’s makes a distinction between naturally private and normatively private situations. According to Tavani (2012: 137), 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.

What is normally considered to be private? This question has also been raised by Weckert and
Adeney (1997) and his answer is that our inner thoughts, our personal relationships, and our personal information such as those that relate to our health and finances are all private matters. Informational privacy is seen in different ways. For example, Islam (2008) and Islam and Bronkovic (2011) consider the exact information on any attribute of an individual (such as the disease diagnosis of a patient, and income of an employee) as private. However, Fule and Roddick (2004) consider any sensitive pattern/trend (such as Client Age $> 35$ & Client Post-Code = “11” $=>$ Loan Approval Risk is High) about a group of individuals as private. While Murahidhar et al. (1999) consider any exact information about a group of individuals as private even if it is not clear which individual the disclosed information belongs to.

Privacy is valued for many reasons including human’s ends like trust and friendships (Tavani, 2012). For Jim Moor, privacy is important because it is the articulation or expression of the core value security (Tavani 2012). 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 (social good) which leads to eroding individual’s freedom. This in turn affects democracy (Johnson 2001). In our view, privacy is also important for love and marriage 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 (Boyd & Ellison 2007, Jones et al 2008, Young 2009, Al-Saggaf 2011). That said, self-disclosure is rampant on both SNS (Jones et al 2008, Valenzuela, Park & Kee 2009) and online communities (Dyson 1998, Horn 1998, Kollock & Smith 1999, Markham 1998, Rheingold 2000). There are many reasons why self-disclosure is common online but the lack of oral and non-verbal cues, and lack of public self-awareness are major facilitators to its

Trust between online communicators has been found to be another factor that encourages self-disclosure (Valenzuela, Park & Kee 2009). Trust is vital for personal relationships (Cocking & Matthews 2000, Weckert 2003); in fact, one way to show the strength of a friendship between two individuals is by demonstrating that they trust other. At the same time to show my trust in someone I have to reveal more about myself. That is why self-disclosure is also important for personal relationships (Preece 2000, Rheingold 2000, Rifkin 2000, Wallace 1999). Indeed, researchers have found that as people become familiar with others online, they tend to reveal more about themselves (Barnes 2001, Horn 1998, Markham 1998).

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. For example 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 the same 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 2012).

**Data Mining Threats to the Privacy of the Users of Facebook and/or other SNS:**

There are various data mining tasks including data collection (such as automatic collection of weather

---

2 The example of women in Saudi Arabia is chosen because the first author has lived in Saudi Arabia for a very long time and has been studying the impact of technology on this conservative society for more than 10 years now.
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. There are a number of pattern extraction techniques including clustering (such as K-means and Fuzzy C-means clustering), classification (such as decision trees and artificial neural networks) and association rule mining such as finding purchase patterns from a market basket dataset.

Data mining can extract various interesting and hidden 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 (2008) and Islam and Bronkovic (2011). For example, the pattern that people from Asia have black eyes is a public knowledge and therefore should not be considered sensitive. If a person is originally from Asia then it is generally assumed that he/she has black eyes and there is nothing wrong with that. However, some general patterns even if when they are applicable on a big number of people can disadvantage an individual and therefore needs to be considered as sensitive. For example, if a bank discovers that all previous loans granted to the customers living in a specific suburb are defaulted and therefore 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 sensitive 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 sensitive to the applicant.

Data Mining can be used for extracting various sensitive patterns of the Facebook or other SNS users. The patterns can then be applied on any Facebook (or any other SNS) users resulting in huge user discomfort and serious breach of individual privacy. For example, a malicious data miner can study the Facebook 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 the 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).

The data miner can also have some supplementary knowledge on his/her friends in addition to the knowledge learned from their Facebook activities. The data miner can also use the supplementary knowledge to create a class attribute of the dataset. The class attribute can be considered as the label of a record. All other attributes are considered as non-class attributes. Classification task (using techniques including decision tree and artificial neural network) builds logic rules from the dataset to classify the records on the values of the class attribute. An example of a logic rule can be “if A = a₁ and B = b₁ then C = c₁,” where A and B are non-class attributes, C is the class attribute and a₁, b₁, c₁ are the attribute values.

Based on his/her supplementary knowledge on each friend in the Facebook, a malicious data miner can add an attribute say “Willingness to date with a male” (Col 7 in Table 1) as the class attribute having values “yes”, “may be” and “no” for the records of the dataset created from the Facebook observation. All other attributes can be considered as non-class attributes. The data miner can then apply a classifier on the dataset and extract sensitive logic rules to classify a Facebook user as either “Willing to date with a Male” or “Not Willing to Date with a Male”. Once the logic rules are built they can be used on any Facebook user (who the data miner even do not know personally) in order to classify as either “Willing to date with a Male” or “Not Willing to Date with a Male” simply based on the users’ Facebook activities. The malicious data miner can then approach the female users who fall in the category “Willing to date with a Male” for a date. This can be disturbing for a Facebook user especially in a conservative society. If a female Facebook user falls in the category “Willing to date with a Male” can have serious social problems including extortion, difficulty in finding a partner and even inferior treatment in a job environment.
We now give a hypothetical example of a dataset, as shown in Table 1, created from Facebook activities and supplementary knowledge of a malicious data miner.

<table>
<thead>
<tr>
<th>Co 1</th>
<th>Co 2</th>
<th>Co 3</th>
<th>Co 4</th>
<th>Co 5</th>
<th>Co 6</th>
<th>Co 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Nationality</td>
<td>Own Pic/week</td>
<td>Exposure</td>
<td>Status/week</td>
<td>Ratio</td>
<td>Willingness</td>
</tr>
<tr>
<td>21</td>
<td>Saudi</td>
<td>7</td>
<td>High</td>
<td>3</td>
<td>0.1</td>
<td>No</td>
</tr>
<tr>
<td>19</td>
<td>Pakistan</td>
<td>2</td>
<td>High</td>
<td>12</td>
<td>5</td>
<td>Yes</td>
</tr>
<tr>
<td>33</td>
<td>Yemen</td>
<td>7</td>
<td>High</td>
<td>2</td>
<td>4</td>
<td>Yes</td>
</tr>
<tr>
<td>35</td>
<td>Saudi</td>
<td>2</td>
<td>High</td>
<td>7</td>
<td>0.1</td>
<td>No</td>
</tr>
<tr>
<td>19</td>
<td>Pakistan</td>
<td>2</td>
<td>Medium</td>
<td>12</td>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>25</td>
<td>Yemen</td>
<td>3</td>
<td>Medium</td>
<td>3</td>
<td>4</td>
<td>No</td>
</tr>
<tr>
<td>22</td>
<td>Saudi</td>
<td>7</td>
<td>Medium</td>
<td>7</td>
<td>0.1</td>
<td>May be</td>
</tr>
<tr>
<td>33</td>
<td>Pakistan</td>
<td>8</td>
<td>Medium</td>
<td>8</td>
<td>5</td>
<td>May be</td>
</tr>
<tr>
<td>21</td>
<td>Saudi</td>
<td>7</td>
<td>Medium</td>
<td>22</td>
<td>5</td>
<td>Yes</td>
</tr>
<tr>
<td>27</td>
<td>Saudi</td>
<td>8</td>
<td>Medium</td>
<td>33</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>22</td>
<td>Saudi</td>
<td>2</td>
<td>Low</td>
<td>22</td>
<td>7</td>
<td>No</td>
</tr>
<tr>
<td>26</td>
<td>Saudi</td>
<td>8</td>
<td>Low</td>
<td>2</td>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>21</td>
<td>Saudi</td>
<td>7</td>
<td>High</td>
<td>3</td>
<td>0.1</td>
<td>No</td>
</tr>
<tr>
<td>19</td>
<td>Pakistan</td>
<td>2</td>
<td>High</td>
<td>12</td>
<td>5</td>
<td>Yes</td>
</tr>
<tr>
<td>33</td>
<td>Yemen</td>
<td>7</td>
<td>High</td>
<td>2</td>
<td>4</td>
<td>Yes</td>
</tr>
<tr>
<td>35</td>
<td>Saudi</td>
<td>2</td>
<td>High</td>
<td>7</td>
<td>0.1</td>
<td>No</td>
</tr>
<tr>
<td>19</td>
<td>Pakistan</td>
<td>2</td>
<td>Medium</td>
<td>12</td>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>25</td>
<td>Yemen</td>
<td>3</td>
<td>Medium</td>
<td>3</td>
<td>4</td>
<td>No</td>
</tr>
<tr>
<td>22</td>
<td>Saudi</td>
<td>7</td>
<td>Medium</td>
<td>7</td>
<td>0.1</td>
<td>May be</td>
</tr>
<tr>
<td>33</td>
<td>Pakistan</td>
<td>8</td>
<td>Medium</td>
<td>8</td>
<td>5</td>
<td>May be</td>
</tr>
<tr>
<td>21</td>
<td>Saudi</td>
<td>7</td>
<td>Medium</td>
<td>22</td>
<td>5</td>
<td>Yes</td>
</tr>
<tr>
<td>27</td>
<td>Saudi</td>
<td>8</td>
<td>Medium</td>
<td>33</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>22</td>
<td>Saudi</td>
<td>2</td>
<td>Low</td>
<td>22</td>
<td>7</td>
<td>No</td>
</tr>
<tr>
<td>26</td>
<td>Saudi</td>
<td>8</td>
<td>Low</td>
<td>2</td>
<td>2</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 1: A sample dataset on Facebook users

A data miner can apply a decision tree algorithm (for example C4.5) on the training dataset shown in Table 1 and build a decision tree as shown in Table 1 that considers “Willingness” as the class attribute. The rectangles denote internal nodes and the ovals denote leaves of the tree. There are six leaves and four nodes in the tree. Each leaf of the tree represents a logic rule that classifies the records belonging to the leaf. For examples, the logic rule for Leaf 1 is “if Exposure=High, and Ratio <= 0.1 then Willingness = No”. There are four records belonging to the leaf meaning that four records have Exposure = High and Ratio <= 0.1 in the training dataset shown in Table 1. In this example, out of the four records all of them have Willingness = No. Therefore, Logic Rule 1 suggests that if a Facebook user uploads pictures with high exposure and the ratio of male friends to female friends is less than or equal to 0.1 then she is usually not willing to date a male. However, Logic Rule 2 (for Leaf 2) suggests that if picture exposure of a Facebook user is High and ratio of male to female friends is
greater than 0.1 then she is likely to be willing to date a male since out of 4 such records in the training dataset all are willing to date a male. Similarly, Leaf 5 and Leaf 6 also represent logic rules that classify records as willing to date.

A malicious data miner can then use the knowledge on any other Facebook users who he/she even does not know. The data miner can learn about the Facebook activities of a user through common friends. Therefore, the data miner may identify a Facebook user as willing to date even though he/she does not know the user. If the logic rules that label a user as willing to date are made public then anyone knowing the rules may classify a user as either willing or not willing to date. This can cause a serious social and emotional trouble for the Facebook users who fall in the category of willing to date especially in a conservative society like Saudi Arabia. This is especially more undesirable for those who are not willing to date but are approached by unwanted people. The logic rules are extracted using data mining algorithms from a sample training dataset and cannot guarantee a correct classification for a new record.

A moral case for the SNS users’ right to privacy

The section titled “Privacy as an Ethical Issue” above discussed some specific reasons why privacy is valued. The aim of this section is to make a moral case for the SNS users’ right to privacy using both
utilitarianism and deontology ethical theories. From a utilitarian perspective, it would appear that using data mining to invade SNS users’ privacy can harm them. For example, malicious SNS users can use data mining technologies to target women on Facebook and threaten them to either submit to their sexual desires or face the consequences of having hidden but sensitive information about them posted on the internet or distributed publically, thereby damaging their family reputation. Putting restrictions on the use of data mining (more on this below) can certainly help stop harm from being inflicted upon those women. This shows that data mining can be used to cause harm to people but that privacy can protect individuals from these kinds of harm.

From a deontological perspective it would appear developers of data mining algorithms have a duty of care to the people who are likely to be affected by the algorithms they develop. While it may be difficult to ensure that these value laden technologies can’t be used by malicious data miners to invade people’s privacy, developers of data mining algorithms have a responsibility to also develop privacy preserving techniques. Respect for persons entails that people should be treated as ends in themselves, and not as means to some end. To treat SNS users with respect means they should be treated as persons who have value in themselves, and not just as pieces of information that could be useful, for example, to third party advertisers. SNS owners should exercise caution when selling/sharing their users’ data to/with third party advertisers so a dataset from users’ data can’t be made available to malicious data miners to mine and use the results against the vulnerable and unsuspecting members of the society and thus breach their privacy.

Privacy protects some freedoms and restricts others. It protects an individual's freedom to be alone, to do things without others intruding and to control what others know about him or her. However, it restricts the extent to which an individual can observe others and what he or she can learn about others (Weckert and Adeney, 1997); although in the case of the SNS, much to the dismay of the users, this may not be possible. That is, privacy may not be able to restrict the freedom of the malicious SNS users to use data mining algorithms to uncover hidden information about a few people with the mission to target them individually because SNS users’ data (necessary for the dataset) and data mining algorithms can easily be acquired. This suggests that privacy in this case may not, when it
should, restrict the freedom of the malicious SNS users to harm others. Thus, while it might be argued that the moral value of privacy comes from its specific function in promoting the liberty of individuals by restricting the scope of the power of those in positions of strength (data miners), in the case of these SNS users, this may not happen (Weckert and Adeney 1997).

But why should the SNS users be concerned about their privacy? The number of times they update their status a week or the level of exposure within the photos they share, for example, are the business of nobody except these SNS users but is this information in itself important to the SNS users? Weckert and Adeney (1997) argue that their preference for unsugared, black coffee rather than the sweet, white variety is also the business of nobody but them and the person making their coffee, but, according to them, worrying about the privacy of this information seems a bit extreme. While much information about the SNS users, like their age, gender, nationality and marital status, which some display on their SNS profiles, might not be much more important than the preference in coffee, the number of times they update their status a week or the level of exposure within the photos they share can reveal a little more about these users when information from other users are also added to the dataset and mined. As mentioned above, a data miner can then apply a data mining algorithm on the dataset to extract rules they can use to classify Facebook users. Once the rules are built they can be used on any Facebook user in order to classify them. The malicious data miner can then approach female users who fall in a particular category. In societies which holds on more than others to their cultural and traditional values and where people compare each other in terms of how observant they are of their cultural and traditional values, receiving threats from strangers to either submit to their sexual desires or face the consequences of having hidden but sensitive information about them posted on the internet or distributed publically does, in effect, subject them to embarrassment and possibly discrimination from their peers if the information about them is out. It could make them feel shameful while encouraging their peers to feel sorry for them or ostracise them or, worst, treat them with disrespect, which are harmful outcomes to these individuals.

**Conclusion and recommendations to mitigate the risk**

This article explored the potential of data mining as a technique that could be used by malicious data
miners to threaten the privacy of SNS users and made a moral case for the protection of users’ privacy. Using a hypothetical dataset, created for the purpose of this article, a data mining algorithm was applied to this hypothetical dataset to demonstrate the ease at which characteristics about the SNS users can be discovered and used in a way that could invade their privacy. This was followed by a short philosophical analysis which argued for the importance of protecting users’ privacy from the threats of data mining.

Privacy is a crucial social good and an instrumental, if not intrinsic, universal value (for the difference between instrumental and intrinsic values see Burmeister, Weckert & Williamson (2011)). Privacy is valued for many reasons (as discussed above). While people generally learn about privacy and how to maintain it from their parents, schools and society, the development of technologies has meant that people are operating in environments that did not exist in the past. As a result, on one hand, people are facing new challenges they have never faced before, on the other hand, they do not seem to get clear guidelines from their parents, schools, and society on how to operate safely on these new environments such as SNS.

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. For example, if the exposure of a user is ‘Medium’ in her pictures then she can deliberately keep the number of her own picture upload per week less than three (3) so that she falls in Leaf 4 (Table 1) and thereby safely classified as not willing to date.

Another way to protect privacy online is by putting 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.

We therefore recommend 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 o the default settings) can be another way to keep malicious data miners out of the way.

The threats from data mining on individuals’ privacy are serious as the application of a data mining algorithm on the hypothetical dataset has shown; and users’ should have a right to their privacy online as the short philosophical analysis has shown. While the recommendations offered here don’t solve the problems, it is hoped the article at least began to raise SNS users’ awareness about the ways in which information that they reveal online can be used by malevolent data miners to harm them and how to operate in SNS safely in the midst of these threats.

References

Al-Saggaf, Y. (2011) “Saudi Females on Facebook: An Ethnographic Study”, International Journal of Emerging Technologies and Society, Faculty of Life and Social Sciences, Swinburne University of Technology, AUS Vol 9 No 1 pp 1-19


Markham, A. N. (1998) Life Online: Researching Real Experience in Virtual Space, AltaMira Publications, Walnut Creek, CA

