DE GRUYTER
OPEN

Research Paper

Twitter Users' Privacy Concerns: What do Their Accounts' First Names Tell Us?

Daniela Fernandez Espinosa & Lu Xiao[†]

School of Information Studies, Syracuse University, New York
Syracuse, New York, 13210

Citation: Daniela Fernandez Espinosa & Lu Xiao (2018). Twitter Users' Privacy Concerns: What do Their Accounts' First Names Tell Us?

Vol. 3 No. 1, 2018
pp 40–53
DOI: 10.2478/jdis-2018-0003

Received: Apr. 19, 2017
Revised: Oct. 4, 2017
Accepted: Nov.16, 2017

Abstract

Purpose: In this paper, we describe how gender recognition on Twitter can be used as an intelligent business tool to determine the privacy concerns among users, and ultimately offer a more personalized service for customers who are more likely to respond positively to targeted advertisements.

Design/methodology/approach: We worked with two different data sets to examine whether Twitter users' gender, inferred from the first name of the account and the profile description, correlates with the privacy setting of the account. We also used a set of features including the inferred gender of Twitter users to develop classifiers that predict user privacy settings.

Findings: We found that the inferred gender of Twitter users correlates with the account's privacy setting. Specifically, females tend to be more privacy concerned than males. Users whose gender cannot be inferred from their provided first names tend to be more privacy concerned. In addition, our classification performance suggests that inferred gender can be used as an indicator of the user's privacy preference.

Research limitations: It is known that not all twitter accounts are real user accounts, and social bots tweet as well. A major limitation of our study is the lack of consideration of social bots in the data. In our study, this implies that at least some percentage of the undefined accounts, that is, accounts that had names non-existent in the name dictionary, are social bots. It will be interesting to explore the privacy setting of social bots in the Twitter space.

Practical implications: Companies are investing large amounts of money in business intelligence tools that allow them to know the preferences of their consumers. Due to the large number of consumers around the world, it is very difficult for companies to have direct communication with each customer to anticipate market changes. For this reason, the social network Twitter has gained relevance as one ideal tool for information extraction. On the other hand, users' privacy preference needs to be considered when companies consider leveraging their publicly available data. This paper suggests that gender recognition of Twitter users, based on Twitter users' provided first names and their profile descriptions, can be used to infer the users' privacy preference.

[†] Corresponding author: Lu Xiao (E-mail: lxiao04@syr.edu).

Originality/value: This study explored a new way of inferring Twitter user's gender, that is, to recognize the user's gender based on the provided first name and the user's profile description. The potential of this information for predicting the user's privacy preference is explored.

Keywords Social media; Twitter; Gender recognition; Privacy preferences

1 Introduction

The decision process for a company must be based on the market preferences, and for this reason, many companies consider that the key to success is to know their customers at a personal level (preferences). A recent study, conducted by IBM, indicates that 48% of consumers want personalized promotions through digital shipments, while 44% expect personalized benefits in physical stores (Gonzalez, 2015).

One critical strategy to renew and stay in the commercial race in today's competitive market is to know how the market is evolving. Customers provoke the evolution of the market, and for this reason companies are investing large amounts of money in business intelligent tools that allow them to predict the preferences of their customers. If companies know the preferences of their customers, they can anticipate changes, assess risks, reduce costs, and increase profits.

According to Howland, CEOs must understand future consumers' needs in order to anticipate trends in the market (Howland, 2014). Companies must invest time in learning about consumer needs, wants, and demands. If they know this, companies will have a "consistent service through strategic dialogue" (Howland, 2014). This will increase consumer loyalty and will attract new potential consumers.

The magazine *Entrepreneur* remarks that market segmentation is the best strategy to gain an understanding of the customers' preferences (Entrepreneur Magazine, 2012). Fran Leo defines market segmentation as the process of dividing the total market of a good or service into several smaller, internally homogeneous groups such as by customers' ages, genders, and lifestyles (Ale, 2015). One of the main objectives of market segmentation is to figure out the preferences of the users in order to provide a more personalized service without the need of conducting direct dialogues. If the segmentation is made properly, the company will reduce the cost of targeted advertising, offer more personalized service, and potentially increase its revenues. For example, previous work has shown that gender plays an important role in the purchase decision-making. Van Aswegen explains that shopping behavior in men and women is highly affected by their considerations, concerns, perspectives, and motives. When companies are developing target ads strategies must recognize gender-specific concerns, to offer what the users really want (van Aswegen, 2015).

On the other hand, today's global market makes it challenging for companies to have direct communication with each customer to know his or her personal preferences, and allocate him/her into one of the identified homogenous groups. For this reason, companies are using Internet users' social media data that allow them to know their customers, without using a form of direct communication and offer them a more personalized service. However, companies face a big challenge in this approach, that is, the Internet users' privacy concern.

For Parker, there is not consensus in the legal and philosophical literature about what is the proper definition of personal privacy (Parker, 1973). Gross and Acquisti said that there are various levels and aspects regarding privacy on social platforms (Gross & Acquisti, 2005). In the social media context, privacy can be understood as the right to control how personal information is collected and used by social platforms.

Although many personal data are freely available on the Internet and easily accessible, this does not mean that the users are not concerned about their privacy. For Greenfield, it is important that companies "don't be creepy using robust user data for ad targeting while respecting privacy" (Greenfield, 2012). Greenfield mentions that there is a fine line between creepy and appropriate. In the online environment, some users are more likely to receive targeted ads and some are not. Greenfield recommends companies to respect the users who are not likely to receive targeted ads because this can cause to lose a customer by a simple ad (Greenfield, 2012). Morey, Forbath, and Schoop (2015) explain the importance of companies to be trusted by their consumers. The authors point out that consumers expressed concern about their privacy. American users rank privacy issues high; 72% prefer not to share personal information with business (Morey, Forbath, & Schoop, 2015). It is important for companies to know their customers in a personal level in order to offer personalized services, but how companies can offer target ads if certain users are not likely to share their personal information. How to identify these users from the pool of Internet users then becomes an important and urgent research question in this direction.

Interested in answering this question, it has been developing computational techniques to help classify Twitter public accounts into privacy-concerned and privacy-not-concerned groups. Twitter has over 330 million active users (Statista, 2016). When an individual creates an account on Twitter, they only must type their complete name, email address, and password.

In 2014, Twitter announced, "advertisers can use Twitter Analytics data to target users more directly and reach the most valuable audiences" (Lopez, 2014). While Twitter users can choose to make their accounts protected, those with public setting can still be privacy concerned—e.g., research has shown that privacy setting in social media is not an accurate indicator of the user's level of privacy concern (Gross & Acquisti, 2005b; Irani et al., 2009).

Compared to other social networks, Twitter works with basic privacy setting: protected or public account. When a user creates an account, his/her information is publicly accessible by default. It is necessary to change the profile settings to make a profile private. Michael Zimmer and Nicholas Proferes argue that not all the users are aware about this default option of privacy, and accidentally can provide more personal information than they intended (Zimmer & Proferes, 2014). For the intentions of this work, we considered users whose profiles are private as users with high levels of privacy concern.

Consumers who go beyond to protect their information in social platforms are not happy about companies knowing their personal information (Lopez, 2014). These consumers' attitudes reveal concerns regarding the privacy use of their data. Prior work has leveraged features from Twitter's tweets and some profile attributes to identify a Twitter public account's level of privacy concern (Khazaei et al. 2016b).

In this paper, we examine whether the first names and profile descriptions in Twitter accounts as clues to infer the user gender provide useful information for predicting the privacy settings of the user. Our work builds on the previous studies that demonstrate the correlation between a person's gender and the level of his/her privacy concern. Many studies indicate that women tend to be more concerned about online privacy and security issues than men (Adam, 2000; Sieger & Moller, 2012). But when the sites are related to shopping or fashion women-related online behaviors are contradictory with their alleged high privacy concerns (Katell, Mishra, & Scaff, 2016; Nazir et al., 2012). Based on 398 online consumer interviews, Riquelme and Roman found that the influence of both privacy and security on online trust was stronger for younger, more educated, and less extroverted males (Riquelme & Román, 2014).

In our work, we use the first name and profile description in a Twitter user account as the main clues to infer the gender of the Twitter user. This approach, that is, to infer the gender in Twitter users through the analysis of the first name, has a reported accuracy of 80.48% (Liu & Ruths, 2013). Previous works have shown that feature combinations of profile attributes increase the accuracy of the result of gender inference (Fernandez, Moctezuma, & Sordia, 2016). We analyzed the profile descriptions through counting the number of feminine and masculine words, and using a Bayesian network with weighted word frequencies to infer the gender of a user. With these results (inferred gender by name, and by profile description), we estimate the gender of a user using Neural Networks algorithm and obtaining 83.47% of accuracy.

Then, we analyzed the correlation between the inferred gender of a user and his/her level of privacy concern. In addition, with the resulting set of features, we create a model to predict user privacy settings. Our best classifier obtained an *F*-score of 0.73, improving in 0.002 points the accuracy of previous classifiers (Khazaei et al., 2016b). This paper is organized as follows. In Section 2, we present some of the related works that are the starting point of our research. In Section 3, we describe our data sets, and process for labeling these data. In Section 4, we explain our methodology for gender recognition. In Section 5, we report our results related to gender and privacy concerns. We summarize our findings and conclude in Section 6. Finally, Section 7 deals with future directions of our work, especially to the accuracy of our method and give detailed advice to companies to help them improve their strategies for market segmentation.

2 Related Work

2.1 Privacy and Gender

Privacy research argues that a user' privacy is affected by many factors such as age, gender, education level, etc. (Adam, 2000). It is also suggested that there are differences between men and women's ethical decision-making in relation to information and computing technologies (ICT's) (Sieger & Moller, 2012). Adam (2000) surveyed a sample of undergraduate and graduate business students and found that women have different responses to men regarding computer privacy. This study shows that women have different views and expectations of privacy in general.

To examine how a person's gender plays a role in his/her perception on privacy, Siegger and Moller (2012) conducted a focus group study, a survey, and an experiment and found that there is a significant difference in the perception of privacy online platforms (applications such as Facebook and Twitter). More specifically, women tend to have higher perception of the security of a particular system or method than men do.

Privacy means different things to different people. Kwasny et al. (2008) tried to understand the privacy views of individuals—young and old, male and female. They conducted a focus group study to investigate folk beliefs about privacy. The participants were 26 students at Georgia Institute of Technology and six older adult females. The main results of this study related to gender-privacy concerns were that females were more likely to talk about privacy involving others (respecting privacy rights and safety) than males. Older adults tended to be more concerned with privacy space than information privacy (Kwasny et al., 2008).

2.2 Analysis and Gender Recognition

A name can reveal many different aspects of a person such as gender, nationality, age, etc. In Twitter, a name gains relevance because it is the first field that a user must fill out in order to create an account.

For inferring the gender in a social platform, specifically in Twitter, Krempeaux recommends looking three main clues on a Twitter profile: the first clue is the profile picture (Krempeaux, 2013). Computer vision techniques can be leveraged to analyze the picture and infer the gender of the user. The second clue is the name, the author remarks that "many given names are highly gender skewed." A simple search in a database of name tags by gender, it could be relatively easy to recognize the gender of the user. Moreover, the last clue is the Twitter profile description. For this last clue, it could be used text analysis for the gender prediction. Krempeaux (2013) explains that tweets can also be used to infer the gender of a user, but the computational cost will be higher than make a simple search on a name database. The accuracy of both methods, tweets text analysis, and the gender prediction strategies (profile, name, and description analysis), may vary, but the result will be substantially constant.

Inspired by Krempeaux work (2013), Herdağdelen published an n-gram data set of Twitter messages, which contains gender of the author and time of posting tags. In order to infer the gender of the author, Herdağdelen used the first name of the users to infer his/her gender, at least, for the ones that provide one. Herdağdelen constructed his own name database based on census statistics. He used the results released by U.S. Census Bureau and U.S. Social Security Administration. His method was simple, if a first name is frequently associated with male users, then all the users who have this first name are classified as male users. The same process is applied for female users. The accuracy of this process was 97.85%, with approximately symmetric error rates for either gender: 2.94% of the females and 2.35% of the males (Herdağdelen, 2013).

Following the same strategy of name analysis to infer the gender, Burger et al. (2011) proposed not only to analyze the name, also, its methodology includes the analysis of the full name, screen name, profile description, and tweet text. The analysis was restrained to only the textual characteristics on the user profile (full name, screen name, profile description, and tweet text). The data set of this experiment contains 4.1 million tweets with 15.6 million distinct features. They used a wide of machine learning and educated guesses to infer the gender including Support Vector Machines, Naïve Bayes, and Balanced Winnow. Balanced Winnow had the highest performance with the accuracy of 74.0%. One of the biggest constraints in this experiment is that the implementation of the classifiers took a long time for training. For example, the implementation of a library for SVM (LIBSVM) got 71.8% accuracy, but required over 15 hours of training.

Fernandez, Moctezuma, and Sordia (2016) proposed a new methodology that analyzes the first name of the user, the profile picture, and the profile description. They worked with a data set of 74,580 users located in Aguascalientes. For the profile picture analysis, they used the API for facial recognition "Face++"[1]. The name analysis was resolved by a binary search into a name database with more than 5,000 names. The name database was constructed based on the INEGI name records of 2010. For the profile description analysis, they made an algorithm that counts the appearance of female or male words. After obtaining the different results for each analysis, they used classification models in order to get the highest accuracy for inferring gender. The highest accuracy (89.94%) was obtained with neural networks and the combination profile picture and name analysis (Fernandez, Moctezuma, & Sordia, 2016).

Liu and Ruths (2013) demonstrated the link between gender and the first name in English tweets. With the 1990 US census as the name data set, they examined the classification power of the Twitter accounts' first names in identifying the users' gender. Their classification results were compared with the human annotation results. The authors showed that for English tweets the first name is highly linked with the gender of the user with an accuracy of 80.48%.

The website HackerFactor Gender Guesser[2] proposes a Bayesian network with weighted word frequencies and parts of speech to estimate the gender of an author. This method has a 60% of accuracy, and makes a distinction between formal and informal writing styles. This approach is a simplified version of the work "Gender Genie" (no longer available), and it uses the same word list and weights (Argamon et al., 2006).

Gender Decoder[3] is a website that counts the number of female and male words in job ads. Its work states that the society uses stereotyped words to refer to people. In this case, they focus on language that is subtly "gender-coded." Gaucher, Fiesen, and Kay (2011) built a list of female and male words, and counted the number of appearances. Based on this result, they show the gender-coded language of the text.

3 Data Set

3.1 Users

For this study, we worked with two different data sets provided by Khazaei et al. (2016a; 2016b). The first data set (data set A) contains 1,645,510 Twitter user accounts from the CNN follower set (this data set does not contain inactive accounts, brands, nor celebrities).

[1] https://www.faceplusplus.com/
[2] http://www.hackerfactor.com/GenderGuesser.php
[3] http://gender-decoder.katmatfield.com/

The second data set has 23,320 user nodes (data set B), and 53 fields per user, including the user's profile attributes, privacy ratio (percentage of protected contacts to all of the contacts of the focal user), the contacts' privacy ratios, etc. To build this data set, a random user was selected by generating a random Twitter ID. After the selection of the first user, the data set was iteratively built in a Breadth First Search (BFS) Manner.

For both data sets, we focused on the profile name and the profile description to infer the gender. The screen name, handle, or alias identifies the user in a short way such as KingJames, and it does not have more than 15 characters. The profile name is provided by the user when he or she creates an account (Twitter, 2017). We concentrated on the field "profile name" because it generally represents the complete name of the user. We put all the names in lower case to facilitate the future searches for a name and infer the gender.

During the gender analysis process, two new columns that indicate the inferred gender were added, one for the inferred gender based on the name analysis, and the second one for the inferred gender based on the profile description. Both columns have three possible values: zero for users that were tagged by females, one for users that were tagged as males, and two for those whose gender was not recognized.

3.2 Name Database

Our name database has two rows; the first row represents the name and the second represents the gender that identifies it. We constructed our database based on the site Behind the Name[④]. Behind the name is a website created by Mike Campbell which helps users to find the meaning of a given name. Multiple users have collaborated with the construction and maintenance of the database. The database displays the name, its usage, history and origin, related names, and popularity around the world (Campbell, 2016).

In order to increase the size of our database we included the names that appeared in the American population census of 1960–2010 (Herdağdelen, 2013) and the most popular names in America in 1990 (Herdağdelen, 2013). Our database has 19,562 unique names (9019 female names, 10,563 male names). These names come from 335 different languages (including fictional languages, such as Elve).

3.3 Training Gender Data Set

Because the name of a user cannot accurately reflect his/her gender, we decided to analyze the profile description of each user as well. We used a data set that

[④] http://www.behindthename.com/

contains 574 Twitter accounts (data set C). These Twitter accounts were manually labeled by human annotators, resulting in 59% male users, 36% female users, and 4% unknown accounts. We added two columns to this data set. The first new column represented the inferred gender based on the name analysis, and the second one represents the gender based on the profile description analysis.

4 Methodology

After constructing the name database, we wrote a Java program to infer the gender of each user in our Twitter user database. We connected our program to the name database to analyze the user names and infer their gender. The program first extracts the complete name of each user and divides it in sub. Then it takes the first substring as the first name and compares it with the names in our name database. It returns the gender if it is found in the name database and tags the user with the gender. If the name is not found in the database, then the user is tagged as "undefined."

Because a name is not enough to infer the gender of a user, we analyzed the profile description. Based on the findings of (Argamon et al., 2006), we wrote a Java program that uses a Bayesian Network with weighted word frequencies and part of speech to estimate the gender of an author. While most of the profile descriptions were in English or Spanish, four users' descriptions were not. We removed them from the data set. Inspired by Gaucher, Friesen, and Kay (2011)'s work, we built our own list of masculine and female words for English and Spanish. It was built based on gender stereotypes and social roles. The female words list contains words such as "wife", "mother", "daughter", "girl", "lovely", "delicate", etc. The male words list contains words such as "husband", "brother", "son", "father", "boy", "player", etc. Our English lists have a total of 67 masculine words, and 75 feminine words. The Spanish version of our list has less number of words because for this language the ending of an adjective determines its gender. We then counted the frequency of masculine and feminine words in the profile description depending on its language and the corresponding gender word list.

After analyzing the name and the profile description, we added two new columns to our three data sets. These columns represented the results of our gender analysis. The training data set (data set C) has three columns that reflect the gender of a user: gender by the human annotators, gender by name search, and gender by profile description analysis. We developed classifiers using this subset of features to obtain the highest accuracy possible regarding to the gender estimation. Table 1 shows the results of the four classifiers. As is shown in the table, Neural Networks classifier had the best performance. We used it to classify the users of our other data sets.

Table 1. Evaluation of classification results.

Algorithm	Precision	Recall	*F*-score
Decision Tree	0.78	0.77	0.78
Support Vector Machine	0.78	0.76	0.77
Neural Networks	0.83	0.82	0.83
Naïve Bayes	0.81	0.82	0.81

We next explored whether and how inferred gender is linked to the privacy settings with data set A. Our first null hypothesis was that the gender of a user, which is reflected from his/her first name and profile description, is independent of the account's privacy setting. We tested our null hypothesis using the chi-square test.

Using the data set A, we analyzed the users whose names do not reveal a specific gender. We were interested in examining whether the fact that a name that does not exist in the name dictionary suggests some level of user awareness of privacy. The data set has one field called HasARealName which represents whether or not any part of the account name can be found in a directory of English names (Khazaei et al., 2016b). The value of this field is one if any part of the account name does appear in the name directory, or zero if it does not. Our second null hypothesis was that whether or not the name exists in the name dictionary is independent of the user's privacy setting.

Lastly, we explored the correlation between a user account's privacy setting and the inferred gender, and then added the inferred gender feature to Khazaei et al. (2016a)'s privacy preference prediction model.

Our results of the hypotheses testing and the performance of the revised model are presented in the results section.

5 Results

At a 0.005 significance level, our chi-square test shows that the gender of a user, reflected from his/her first name and the profile description, is correlated with the user account's privacy setting ($P < 0.0001$, $\chi^2 = 14075.06$). The Cramer's V measure of our chi-square test's effect size is 0.0089 (DF = 3), which is considered to be a small effect (Cohen, 1988). As shown in Table 2, female users tend to be protected users. This is consistent with the findings from the related literature. For example, Adam (2000) states that women are more aware of their personal protection because they consider privacy issues more than men do.

Our examination of the undefined users shows that they tend to use names that do not exist in the name dictionary and have protected accounts. Table 3 shows undefined protected users against undefined public users and the use of a name that

Table 2. User accounts' privacy setting (protected *vs* public) and their gender inferred from their names and profile descriptions.

	Female	Male	Undefined	Total
Protected Accounts	227,238 (55.34%)	288,235 (44.68%)	268,915 (45.38%)	784,388 (47.66%)
Public Accounts	183,358 (44.65%)	354,166 (55.13%)	323,594 (54.61%)	861,118 (52.33%)
Total	410,596 (100%)	642,401 (100%)	592,509 (100%)	1,645,506 (100%)

exists or non-exists. At 0.005 significance level, we rejected the second null hypothesis (p < 0.0001). This result indicates that users whose gender cannot be inferred from the names tend to be aware of the privacy setting.

Table 3. Users whose genders cannot be inferred from the names and the privacy setting of their accounts.

	HasARealName (true)	NotHasARealName	Total
Protected	106,053 (55.37%)	217,541 (54.24%)	323,594 (54.61%)
Public	85,451 (44.62%)	183,464 (45.75%)	268,915 (45.38%)
Total	191,504 (100%)	401,005 (100%)	592,509 (100%)

In the last task of identifying the predictive power of the inferred gender for the account's privacy setting, we considered two categories for the inferred gender: RevealedGender that includes users identified as females or males, and or NotRevealedGender that includes only the users tagged as undefined. Our Spearman correlation test showed that there is a negative correlation between the inferred gender and the account's privacy ratio (r = -0.113, p < 0.0005), which means that the higher the privacy ratio (i.e., the more privacy concerned) the less likely that the gender can be inferred from the names and the profile descriptions.

Khazaei et al. (2016a) proposed a model that used data set A to predict a Twitter user's privacy setting based on the privacy ratios of various profile attributes and on the linguistic features of the profile descriptions. With this data set, we added gender as an additional feature and examined the added predictive power. The comparison of the results is provided in Table 4. As is shown in the table, the addition of the inferred gender feature improved the prediction performance slightly by about 2%.

Table 4. Comparison of our classification results with Khazaei et al. (2016a).

Results obtained by Khazaei et al. (2016a) classifiers				Results obtained by adding gender		
Algorithm	Precision	Recall	*F*-Score	Precision	Recall	*F*-Score
Naïve Bayes	0.66	0.67	0.66	0.67	0.68	0.66
Regression	**0.71**	**0.70**	**0.71**	**0.73**	**0.72**	**0.73**
Logistic	0.69	0.70	0.69	0.69	0.71	0.69
J48	0.68	0.66	0.67	0.67	0.68	0.67
KNN	0.67	0.59	0.63	0.67	0.59	0.63

6 Discussion and Conclusion

It is known that not all twitter accounts are real user accounts, and social bots tweet as well (Chu et al., 2010). A major limitation of our study is the lack of consideration of social bots in the data. In our study, this implies that at least some percentage of the undefined accounts, that is, accounts that had names non-existent in the name dictionary, are social bots. It will be interesting to explore the privacy setting of social bots in the Twitter space.

Nonetheless, our study suggests that companies that are conscientious of the various levels of customers' privacy concerns may leverage the provided names and profile descriptions in Twitter accounts to infer the accounts' privacy concern.

Author Contributions

L. Xiao (lxiao04@syr.edu, corresponding author) devised the project, the main conceptual ideas, and proof outline; instructed the research team, and wrote and revised the manuscript. D. Fernandez (dferna01@syr.edu) performed the experiments, derived the models, and analyzed the data, and wrote the manuscript in consultation with L. Xiao.

References

Adam, A. (2000). Gender and computer ethics. ACM SIGCAS Computers and Society, 30(4), 17–24.

Ale, F.L. (2015). What is market segmentation? (Merca 2.0) Retrieved from http://www.merca20.com/que-es-la-segmentacion-de-mercados/.

Argamon, S., Koppel, M., Fine, J., & Shimoni, A.R. (2006). Gender, genre, and writing style in formal written texts. Text-Interdisciplinary Journal for the Study of Discourse, 23(3), 321–346.

Burger, J.D., Henderson, J., Kim, G., & Zarrella, G. (2011). Discriminating gender on Twitter. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (pp. 1301–1309). Stroudsburg, PA: Association for Computational Linguistics.

Campbell, M. (2016). About our site. (Behind the Name: the Etymology and History of First Names). Retrieved from http://www.behindthename.com.

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2010). Who is tweeting on Twitter: Human, bot, or cyborg? In Proceedings of the 26th Annual Computer Security Applications Conference (pp. 21–30). New York: ACM.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Entrepreneur Magazine. (2012). How to know your market. Retrieved from https://www.entrepreneur.com/article/264931.

Fernandez, D., Moctezuma, D., & Sordia, O. (2016). Features combination for gender recognition on Twitter Users. IEEE International Autumn Meeting on Power, Electronics and Computing, 17(17), 400–408.

Gaucher, D., Friesen, J., & Kay, A.C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. Journal of Personality and Social Psychology, 101(1), 109–128.

Gonzalez, F. (2015). New Twitter tool promising better targeting. (Merka 2.0). Retrieved from http://www.merca20.com/nueva-herramienta-de-twitter-que-promete-una-mejor-segmentacion-del-target/.

Greenfield, C. (2012). Don't be creepy: Using robust user data for ad targeting while respecting privacy. (Target Marketing). Retrieved from http://www.targetmarketingmag.com/post/don-t-creeper-utilizing-robust-user-data-ad-targeting-while-respecting-privacy/all/.

Gross, R., & Acquisti, A. (2005). Information revelation and privacy in online social networks. In Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society (pp. 71–80). Alexandria: ACM.

Herdağdelen, A. (2013). Twitter n-gram corpus with demographic metadata. Language Resources and Evaluation, 47(4), 1127–1147.

Howland, S. (2014). Anticipate trends ensure business success. (Chiefexecutive.net) Retrieved from http://chiefexecutive.net/anticipate-trends-ensure-business-success/3/.

Irani, D., Webb, S., Li, K., & Pu, C. (2009). Large Online Social Footprints–An Emerging Threat. In CSE '09 Proceedings of the 2009 International Conference on Computational Science and Engineering (pp. 271–276). Washington, D.C.: IEEE Computer Society.

Katell, M.A., Mishra, S.R., & Scaff, L. (2016). A fair exchange: Exploring how online privacy is valued. In the 49th Hawaii International Conference on System Sciences (HICSS) (2016) (pp: 1881–1890). Washington, D.C.: IEEE Computer Society.

Khazaei, T., Xiao, L., Mercer, R.E., & Khan, A. (2016a). Privacy Preference Inference via Collaborative Filtering. In Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM), 611–614.

Khazaei, T., Xiao, L., Mercer, R.E., & Khan, A. (2016b). Privacy behaviour and profile configuration in Twitter. In Proceedings of the 25th International Conference Companion on World Wide Web (pp. 575–580). Montreal.

Krempeaux, C.I. (2013). Predicting gender on Twitter. (Charles Iliya Krempeaux Personal Site) Retrieved from http://changelog.ca/log/2013/03/03/twitter_gender.

Kwasny, M., Caine, K., Rogers, W.A., & Fisk, A.D. (2008). Privacy and technology: Folk definitions and perspectives. In CHI '08 Extended Abstracts on Human Factors in Computing Systems (pp. 3291–3296). New York: ACM.

Liu, W., & Ruths, D. (2013). What's in a name? Using first names as features for gender inference in Twitter. In Analyzing Microtext: Papers from the 2013 AAAI Spring Symposium (pp. 10–16). Palo Alto, CA: AAAI Press.

Lopez, N. (2014). Twitter advertisers can now target ads based on the apps a user has installed. (The next web). Retrieved from http://thenextweb.com/insider/2014/12/09/twitter-advertisers-can-now-monitor-user-behavior-mobile-apps-ad-targeting/.

Morey, T., Forbath, T., & Schoop, A. (2015). Customer data: Designing for transparency and trust. (Harvard Business Review). Retrieved from https://hbr.org/2015/05/customer-data-designing-for-transparency-and-trust.

Nazir, S., Tayyab, A., Sajid, A., Rashid, H.u., & Javed, I. (2012). How online shopping is affecting consumers buying behavior in Pakistan? IJCSI International Journal of Computer Science Issues, 9(3), 486–495.

Parker, R.B. (1974). A definition of privacy. Rutgers Law Review, 27(1), 275–296.

Riquelme, I.P., & Román, S. (2014). Is the influence of privacy and security on online trust the same for all type of consumers? Electronic Markets, 24(2), 135–149.

Sieger, H., & Moller, S. (2012). Gender differences in the perception of security of mobile phones. In Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services Companion (pp. 107–112). New York: ACM.

Statista. (2016). Global social networks ranked by number of users. Retrieved from https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

Twitter. (2017). Users. Retrieved from https://dev.twitter.com/overview/api/users.

van Aswegen, A. (2015). Women vs. men—Gender differences in purchase decision making (Guided Selling) Retrieved from http://www.guided-selling.org/women-vs-men-gender-differences-in-purchase-decision-making/.

Zimmer, M., & Proferes, N. (2014). A topology of Twitter research: Disciplines, methods, and ethics. Aslib Journal of Information Management, 66(3), 250–261.

This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (http://creativecommons.org/licenses/by-nc-nd/4.0/).