Perceived Versus Actual Predictability of Personal Information in Social Networks

Conference Paper · September 2016

DOI: 10.1007/978-3-319-45982-0_13

5 authors, including:

Eleftherios Spyromitros-Xioufis
Expedia
31 PUBLICATIONS **2,002** CITATIONS
SEE PROFILE

Georgios Petkos
Information Technologies Institute (ITI)
24 PUBLICATIONS **724** CITATIONS
SEE PROFILE

Symeon Papadopoulos
The Centre for Research and Technology, Hellas
256 PUBLICATIONS **4,720** CITATIONS
SEE PROFILE

Rob Heyman
Vrije Universiteit Brussel
26 PUBLICATIONS **180** CITATIONS
SEE PROFILE

Perceived versus Actual Predictability of Personal Information in Social Networks

Eleftherios Spyromitros-Xioufis[1], Georgios Petkos[1], Symeon Papadopoulos[1], Rob Heyman[2], and Yiannis Kompatsiaris[1]

[1]CERTH-ITI, Thessaloniki, Greece
{espyromi,gpetkos,papadop,ikom}@iti.gr
[2]iMinds-SMIT, Vrije Universiteit Brussel, Brussels, Belgium
rob.heyman@vub.ac.be

Abstract. This paper looks at the problem of privacy in the context of Online Social Networks (OSNs). In particular, it examines the predictability of different types of personal information based on OSN data and compares it to the perceptions of users about the disclosure of their information. To this end, a real life dataset is composed. This consists of the Facebook data (images, posts and likes) of 170 people along with their replies to a survey that addresses both their personal information, as well as their perceptions about the sensitivity and the predictability of different types of information. Importantly, we evaluate several learning techniques for the prediction of user attributes based on their OSN data. Our analysis shows that the perceptions of users with respect to the disclosure of specific types of information are often incorrect. For instance, it appears that the predictability of their political beliefs and employment status is higher than they tend to believe. Interestingly, it also appears that information that is characterized by users as more sensitive, is actually more easily predictable than users think, and vice versa (i.e. information that is characterized as relatively less sensitive is less easily predictable than users might have thought).

Keywords: privacy, social networks, personal attributes, inference

1 Introduction

Online Social Networks (OSNs) have had transforming impact on the overall Internet landscape. OSNs have affected the way people communicate, are being informed or even make business online. An issue that is sometimes overlooked though is the exposure of personal information through the OSNs. Participation in an OSN means that a certain amount of data related to the user is accessible from a) other OSN users and b) the OSN service. The disclosure of specific types of information may pose serious threats to the users. For instance, in several cases, information about the gender, age, ethnicity, political or religious beliefs, sexual preferences, and financial status of a person have been used for unjustified discrimination, for instance, in the context of personnel selection [2] and for loan approval and pricing based on social media profiles [24].

In this paper we look into the issue of privacy in the context of OSNs. In particular, we study the predictability of various types of personal information based on shared OSN data, and contrast it to the users' perceptions about the exposure of their personal information. To perform this analysis, we employ a real life dataset that was composed through a user study that involved 170 participants. Each participant was asked to answer a questionnaire that included questions about his/her personal information as well as questions about his/her perceptions with respect to the disclosure of different types of information, Moreover, all users granted us access to their Facebook data (posts, likes and images) via a specially designed Facebook application.

Utilizing the collected OSN data and user responses, we train and evaluate the accuracy of classifiers that predict various personal user attributes using the OSN data as input. Different classifiers and a number of meta-learning techniques are tested (such as fusion of different feature modalities and multi-label classification). Eventually, we obtain indications of the actual predictability of different types of personal information and compare them to users' perceptions about the predictability and sensitivity of the corresponding types of information. It appears that users' perceptions about the predictability of different types of information are sometimes correct and sometimes not. For instance, users tend to correctly believe that their demographics information, such as their age, gender and nationality can be predicted quite accurately, whereas they incorrectly believe that their political beliefs cannot be accurately predicted. Moreover, it appears that information that is characterized by users as more sensitive is actually more easily predictable than users might have thought, and vice versa. To the best of our knowledge, this is the first study to compare users' perceptions about the disclosure of their personal information through an OSN to the actual predictability of such information.

2 Background

2.1 Privacy in OSNs

The current social research on privacy in OSNs focuses on awareness, attitudes and practices [9] with regard to volunteered or observed personal information disclosure. Nevertheless, it neglects to explore awareness or attitudes with regard to *inferred* information, the third category of personal information identified by the World Economic Forum [11], which is the type of information we focus on in this paper. Several studies have investigated the attitudes of people towards information disclosure in OSNs. For instance, [17] identifies three main classes of users with respect to the level of information disclosure in OSNs: a) privacy fundamentalists, b) pragmatists, and c) unconcerned. Other studies compare attitude with behavior; these studies map what users are willing to disclose and how this is reflected in settings and other proxies that reflect their behavior [6]. A related study [19] shows that OSN users have difficulties dealing with privacy within OSNs. In particular, among 65 users that were asked to look for sharing violations in their OSN profiles (i.e. cases in which they shared content with

people that they really would not like to) every one found at least one such violation. This mismatch between intended and actual sharing policies has been attributed to incomplete information, bounded cognitive ability and cognitive and behavioral biases [1], which may be caused by the difficulty of managing privacy settings and opting-out defaults [15]. While few works have studied privacy with regard to observed data by first [31] and third parties [3], to the best of our knowledge this is the first work to investigate awareness, attitudes and behavior with regard to inferred information on OSNs.

This line of research is significant for two reasons. First, the existence and use of inferred data will increase. Secondly, on a theoretical level, little sociological or psychological models exist that take inferred information into account for privacy. Behavioural economics [1], Westin's [34] privacy definition, contextual integrity [20] and Communication Privacy Management (CPM) [23] are all limited to access control. This means that each privacy perspective presents privacy as a question of giving access or communicating personal information to a particular party. This is illustrated in Westin's definition of privacy: "The claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others." [34]. But for inferred information, this definition becomes: "The claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is *inferred*." However, such control is non-existent because users: a) are unaware and b) have no control over the logic of the inferences being made. Since this area is under-researched, our first aim is to understand if and how users intuitively grasp what can be inferred from their disclosed data.

2.2 Prediction of Personal Attributes

A major issue about privacy is the fact that information about a user may not only appear in an explicit manner, but it can also appear implicitly and may be obtained using appropriate inference mechanisms. For instance, one might easily guess that a user who is interested in university/educational issues is very likely to be a young adult. In the following, we briefly review some previous work on inferring personal information based on OSN data.

In the study of Kosinski et al. [18], 58,466 Facebook users provided their complete like history (170 likes/person on average), their profile information, as well as the results of several psychometric tests. Using likes data, and particularly a reduced (via Singular Value Decomposition) version of the user-like matrix as input, the authors trained linear and logistic regression models to predict numeric and binary variables respectively. The Area Under ROC Curve (AUC) scores for predicting the binary variables were: 95% for ethnicity, 93% for gender, 88% for gays, 75% for lesbians, 85% for political affiliation, 82% for religion, 73% for cigarette smoking, 70% for alcohol consumption, 67% for relationship status, 65% for drug use and 60% for parents being together when the user was 21. The Pearson correlation coefficient for age was 0.75.

Schwartz et al. [28] studied a dataset of 15.4 million status updates from a total of 74,941 Facebook users, who also submitted their gender, age and Big-5

personality scores. They tested traditional techniques of linking language with personality, gender and age such as the Linguistic Inquiry and Word Count (LIWC), which uses a lexicon with pre-selected categories, but also developed a new approach, the Differential Language Analysis (DLA), which generates the lexicon categories based on the text being analyzed. The researchers first used Principal Component Analysis (PCA) to reduce the feature dimension, and then a linear Support Vector Machine (SVM) for classifying gender and ridge regression for predicting age and personality traits. Among other results, they were able to predict gender with 92% accuracy.

Other approaches have looked at a variety of user attributes and techniques. For instance, Backstrom and Kleinberg [4] managed to predict whether a user is single or not with 68% accuracy and whether he/she is single or married with 79% accuracy. Jernigan et al. [16] looked at sexual orientation and achieved an accuracy of 78%. Of particular interest is the study presented by Zheleva and Getoor [35], where different OSNs were considered; examined user attributes are the country, gender and political views. Rao et al. [25] evaluated the accuracy of predicting gender (72%), age (74%), regional origin (77%) and political affiliation (83%) from Twitter messages. Particularly good results (95% accuracy) on political views were obtained by Conover et al. [8]. Very good results on political views (89% accuracy) were also achieved by Penna et al. [21]. Interestingly, they utilized a set of network attributes as features, whereas they also consider two more attributes: for ethnicity they achieved an F-score of 70% and for predicting whether a person is a Starbacks fan an F-score of 76%. Finally, an interesting finding is that inferring personal information based on OSN data can be highly unreliable (close-to-random) for a significant number of users [32].

3 Data Collection and Experimental Setup

3.1 Data Collection

Our study is based on a set of 170 Facebook users who gave us their informed consent to collect their OSN data through a test Facebook application[1]. In particular we collected each user's likes, textual posts and images. In addition, all users answered a questionnaire that included questions about several personal attributes as well as questions related to their perceptions about the predictability and the sensitivity of different types of information. Feedback about the perceived predictability was provided by the users with a yes/no answer to the question: "Can this particular type of information be inferred based on your OSN data?", and feedback about the sensitivity of different types of information was provided in a scale from 1 to 7 with higher values denoting higher sensitivity. Personal user attributes were organized into 10 categories: 1. "Demographics", 2. "Employment status and income", 3. "Relationship status and living condition", 4. "Religious views", 5. "Personality traits", 6. "Sexual orientation", 7. "Political attitude", 8. "Health factors", 9. "Location", 10. "Consumer profile", hereafter

[1] `https://databait.hwcomms.com`

referred to as *disclosure dimensions* [22], to facilitate a more compact and intuitive presentation and handling of a user's personal information. For instance, the "Demographics" dimension includes the following personal attributes: "age", "gender", "education level", "language", "nationality" and "residence". Due to space limitations, the full organization of personal user attributes into disclosure dimensions along with some statistics about the collected data are provided in a supplementary document[2].

3.2 Experimental Setup

In the learning experiments we considered 96 questions from the questionnaire, corresponding to 9 of the 10 disclosure dimensions (location was not considered due to the high cardinality of possible responses, which would lead to a very sparse training set given the limited number of test users). Evaluation was performed using repeated random sub-sampling validation. In this procedure, the data is randomly split n times into training and test sets. For each split, a model is fit to the training set and its prediction accuracy is assessed on the test set. The final performance is calculated as the average over the n tests. For this study, 66% of the data were used for training and the process was repeated 10 times. Since for many of the questions (user attributes), the distribution of responses is highly imbalanced we used AUC as the evaluation measure due to its better robustness with imbalanced classes compared to measures such as classification accuracy that tend to favor classifiers that frequently predict the majority class.

The features that we extract from the OSN data and use as input attributes for the classification models throughout the experiments are the following:

- `likes`: A binary vector where each variable indicates the presence or absence of a like in the set of likes of the user. The vocabulary consists of the 3,622 likes that appear in the sets of likes of at least two users.
- `likesCats`: Each like in Facebook is assigned to a general category, such as "Community" or "Music". This vector is a histogram of the frequencies of these categories in the set of likes of each user. The vocabulary consists of the 191 categories that appear in the sets of likes of at least two users.
- `likesTerms`: A Bag-of-Words (BoW) vector computed using the terms that appear in the description, title and about sections of all likes made by each user. We performed stop-word removal (using three language-specific lists of stop words for the three main languages that appear in the collected content: English, Dutch and Swedish) and kept only terms that appear in the profiles of at least two users. This resulted in a vocabulary of 62,547 terms.
- `msgTerms`: A BoW vector computed using the terms that appear in all posts of each user. The same pre-processing was applied as in the case of `likesTerms`, resulting in a vocabulary of 24,990 terms.
- `LDA-t`: The distribution of topics in the textual content of the user's posts and likes (description, title and about sections). Topics were extracted using

[2] `http://usemp-mklab.iti.gr/usemp/prepilot_survey_data_statistics.pdf`

Latent Dirichlet Allocation (LDA) [5] and different setups involving different numbers of topics were examined (t = 20, 30, 50 and 100 topics).
- `visual`: The concepts depicted in the images of the users. These concepts were detected using the Convolutional Neural Network (CNN) variant presented in [13]. For each image the 12 most dominant concepts were kept, which resulted in a vocabulary of 11,866 distinct concepts for the whole collection. We used three alternative representations:
  - `visual-bin`: a binary vector representing concept presence/absence.
  - `visual-freq`: a histogram vector representing concept frequencies.
  - `visual-conf`: a vector where each variable represents the sum of detection scores for each concept across all images of each user.

4 Experimental Results

Here, we first present a set of thorough learning experiments with the goal of assessing the predictability of different types of user attributes and then compare these results to the perceptions of users.

The first experiment explores the performance of various baseline and state-of-the-art classifiers using the features described in the previous section. In particular, the following classifiers were considered:

- `knn`: The k-nearest neighbors ($k = 10$) classifier using the Euclidean distance.
- `tree`: A simple decision tree classifier (Weka's `REPTree` class [14]).
- `nb`: The Naïve Bayes classifier.
- `adaboost`: The Adaboost M1 boosting meta-classifier with a decision stump (a one-level decision tree) as the base classifier [12].
- `rf`: The Random Forest classifier [7] using 100 random trees.
- `logistic`: An efficient implementation of L2-regularized logistic regression from LibLinear [10] with probabilistic estimates and tuning of the regularization parameter ($c \in \{0.1, 1, 10\}$).

Due to the high computational cost of evaluating all six classifiers using all types of features, instead of performing the evaluation on all 96 target attributes, we selected eight representative ones: 'BMI class', 'Income', 'Health', 'Use of cannabis', 'Smoking behavior', 'Employment status', 'Drinking behavior' and 'Sexual orientation'. For each classifier, Figure 1 shows the best achieved AUC performance (across all types of features) on each target attribute. We see that `logistic` and `rf` are the two best-performing classifiers in most cases. Specifically, `logistic` achieves the best performance in five targets, `rf` in two targets and `adaboost` in one target. Given the good performance of `logistic` and `rf` and their better scalability (especially with respect to the number of features) compared to the competing classifiers, we opted for using these two classifiers in the rest of the experiments.

Our next experiment aims at evaluating the relative strength of the different types of features described in Section 3.2. Figure 2 shows the average AUC performance (across all 96 target attributes) using each type of feature by each

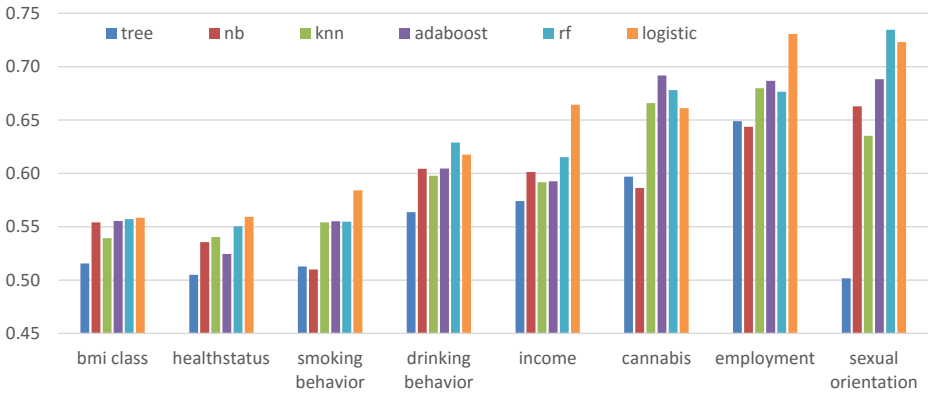

Fig. 1. Comparison of the performance of six different classifiers on eight target attributes. The best performance across all types of features is reported.

classifier (`logistic` and `rf`). We observe that the best performance is obtained with `likes`, followed by `LDA-t`, `likesTerms`, `msgTerms` and `likesCats`. On the other hand, features based on visual concepts obtain lower performance scores, indicating that it is difficult to predict user attributes using this type of information alone. `LDA-30` has a small edge over other LDA-based features, while `visual-conf` obtains the best performance among features based on visual concepts. With respect to the two classifiers, `logistic` is consistently better (on average) than `rf` with all feature types.

Since different features may capture different information about users, we also explore the possibility of increasing performance by combining features. To this end, we employ a simple late fusion scheme that consists of averaging the results produced by different single-feature classifiers. In this experiment, we use only the `logistic` classifier (as it was shown to significantly outperform `rf` in the previous experiment) and evaluate the performance of all possible two-classifier combinations. To avoid combining features that carry redundant information, we selected only the best performing variants of `LDA-t` (`LDA-30`) and `visual` (`visual-conf`). Thus, we ended up evaluating all 15 distinct pairs of the following features: `likes`, `likesTerms`, `likesCats`, `msgTerms`, `LDA-30`, and `visual-conf`. Figure 3 shows the average performance obtained by different late fusion schemes, along with the performance of models that are based on single features to facilitate a direct comparison. We see that the top four late fusion schemes include `LDA-30` features and that two of them, `LDA-30/likes` and `LDA-30/visual-conf` obtain slightly better performance than the performance of the best single-feature model (the one based on `likes`). Another interesting observation is that although `visual-conf` and `likesCats` are the two worst performing features when used separately, their combination with `LDA-30` provides better results compared to e.g. the combination of `LDA-30` with `msgTerms`. This is attributed to the fact that `msgTerms` are computed from the same data

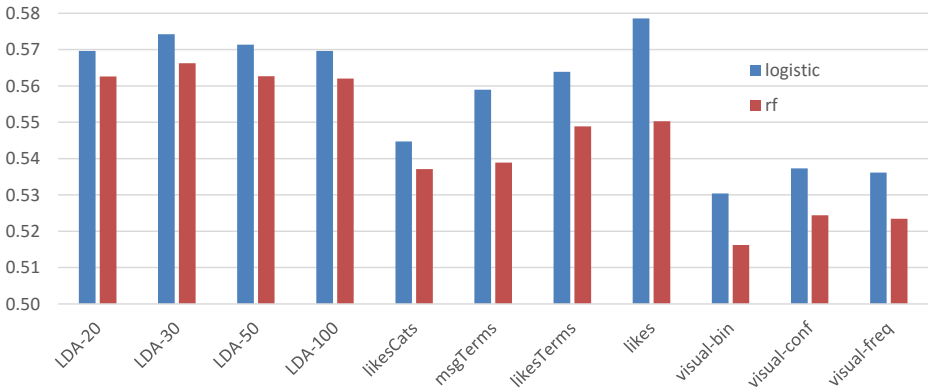

Fig. 2. Average AUC (across all 96 classification targets) for each type of feature using `logistic` and `rf`.

(terms appearing in a users likes) as the `LDA-30` features and thus exhibit a lower degree of complementarity with them compared to `likeCats` and especially `visual-conf` features.

In classification problems where multiple target variables need to be predicted based on a common set of predictive variables, predictive performance can often be improved by taking target correlations into account [33]. Recognizing that different user attributes are likely correlated, we studied whether we could further improve predictive performance using *multi-label classification* methods. However, differently from typical multi-label classification problems where all target variables are binary, here we deal with a more general learning task since in addition to binary variables we also have to predict nominal variables with more than two levels. As a result, multi-label classification approaches that transform the problem into one or more multi-class classification problems where each class corresponds to a different combination of labels (e.g. [26]) are not directly applicable. On the other hand, approaches that build a separate model for each target can be easily adapted to handle different types of target variables by employing appropriate base models as shown in [30].

This category includes the baseline Single-Target (ST) approach that builds an independent model for each target variable and does not account for target dependencies (the approach that we have used so far), but also approaches that capture target dependencies by treating other target variables as additional feature attributes when predicting each target. A popular approach of this type is Classifier Chains (CC) [27]. CC constructs a chain of models, where each model involves the prediction of a single target and is built using a feature space that is augmented by the targets that appear earlier in the chain. During prediction, where the target values are unknown, CC uses estimated values obtained by sequentially applying the trained models. Here, we use this approach to predict a mixture of binary and nominal target variables by employing a multi-class

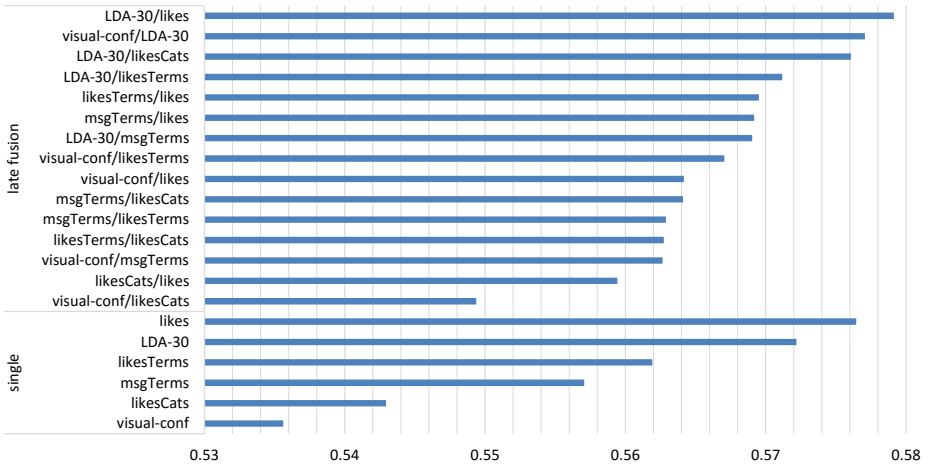

Fig. 3. Average AUC (over all 96 classification targets) of single-feature models and of models that combine two features with late fusion.

instead of a binary classifier for nominal target variables with more than two levels. In addition to CC, we also use an ensemble version of the method called Ensemble of Classifier Chains (ECC) [27]. ECC builds multiple differently ordered random chains of classifiers and the final prediction for each target comes from majority voting.

We evaluated ST, CC and ECC (using 10 random chains) on each of the 96 targets using `likes` and `LDA-30` features (the two best performing features). All methods take the base single-target classifier as a parameter. Thus, we instantiated each method with `logistic` and `rf` and report, for each target, the best performance obtained using any combination of base classifier and feature. Figure 4 shows the results obtained by each method on the 28 targets related to the "Consumer profile" dimension (we do not show results on all 96 targets to improve the readability of the figure). We see that, although ST obtains the best performance in most targets (17 out of 28), it is outperformed by CC in 2 targets and by ECC in 9 targets. The picture is similar when all targets are considered. Again, despite the fact that ST obtains the best performance in most targets (50 out of 96), it is outperformed by CC in 17 targets and by ECC in 29 targets. As expected, ECC outperforms CC in most cases. A closer look at the results, reveals that CC tends to perform better than ST on targets that appear earlier in the chain but the performance starts deteriorating after a certain number of targets. This is due to the fact that prediction noise is accumulated along the chain, a known problem for CC on datasets with many targets such as this one.

Figure 5 shows the best AUC that we obtained for each target attribute, using any combination of features and classification approach. Target attributes are grouped by disclosure dimension and sorted in ascending AUC order within each dimension. The average best AUC achieved for all 96 attributes is 0.63

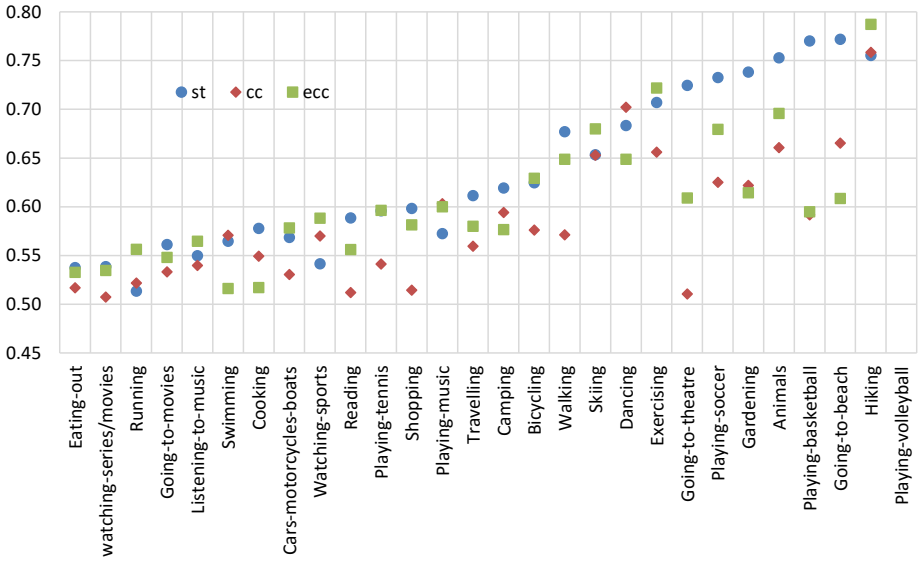

Fig. 4. Maximum AUC per target using ST, CC and ECC.

which represents a significant improvement over random performance (0.5) and is actually quite impressive if we take into account the limited number of training examples and the high cardinality of some classes.

Having performed a set of thorough experiments that measured the actual predictability of different types of personal information, we now proceed to a comparison with the perceived predictability and sensitivity of different types of information (according to users' responses in the survey). Table 1 presents the ranking of dimensions according to a) their perceived predictability, b) their actual predictability according to our experiments (obtained by averaging the performance over the attributes of each dimension) and, c) their predictability according to [18] (for those dimensions for which data is available). It is noted that users perceive "Demographics" as the dimension that is most predictable (88.4%), and indeed it was found through our study that it is the dimension that can be predicted most accurately. Our conclusions also appear to mostly match those of [18]. In particular, "Demographics" and "Political views" are identified as the most predictable dimensions in both studies and the ranking of the remaining dimensions is quite similar (except for "Religious views").

Figure 6 presents an overall comparison between perceived and actual predictability of dimensions with respect to perceived sensitivity. Let us first focus on the relationship between perceived predictability and sensitivity. With the exception of the "Religious views" and "Relationships" dimensions, there appears to be a clear linear relationship between sensitivity and perceived predictability. That is, the more sensitive some dimension is perceived by users, the less predictable it is considered. For instance, "Demographics", the dimension that

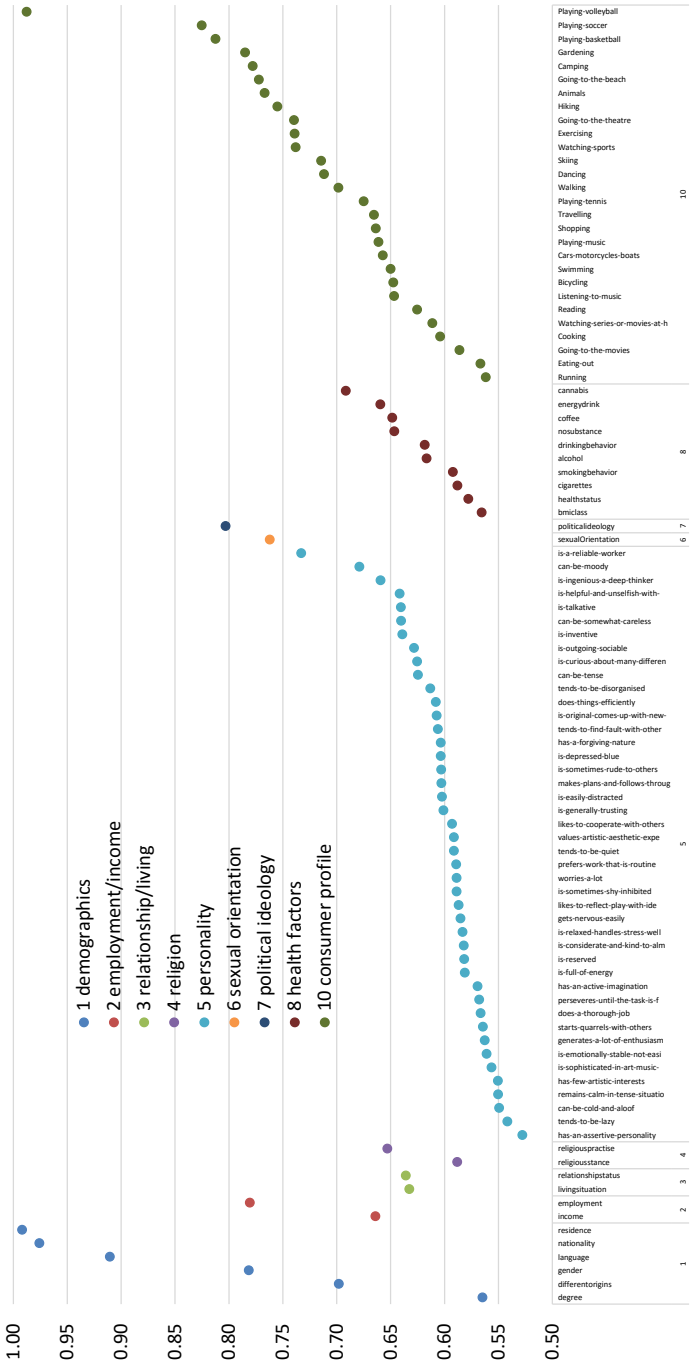

Fig. 5. Best AUC achieved on each target attribute using any combination of features, classifier and fusion approach.

Table 1. Ranking of dimensions according to a) perceived predictability, b) actual predictability (according to our study) and, c) actual predictability according to [18].

Rank	Perceived predictability of dimension	Actual predictability according to our study	Actual predictability according to [18]
1	Demographics	Demographics	Demographics
2	Location	Political views (**+4**)	Political views
3	Relationship status and living condition	Sexual orientation	Religious views
4	Sexual orientation	Employment/Income (**+5**)	Sexual orientation
5	Consumer profile	Consumer profile	Health status
6	Political views	Relationship status and living condition	Relationship status and living condition
7	Personality traits	Religious views (**+1**)	
8	Religious views	Health status (**+1**)	
9	Employment/Income	Personality traits	
10	Health status		

is perceived as the easiest to predict (and is actually the most predictable), is considered to be the least sensitive. At the same time, "Health status" the dimension that is perceived as the least predictable (and is actually among those that are the hardest to predict), is considered as the most sensitive.

Two more observations can be made based on the results shown on Figure 6. The first is that the accuracy of the perceptions of users about the predictability of each dimension tends to vary considerably. For some dimensions, their perception is rather accurate, but for others it is far from accurate. For instance, users correctly believe that their demographics information is quite predictable (actual predictability is quite high) and also have a quite accurate perception about the predictability of their consumer profile information and factors related to their personality traits. On the other hand, their perception about the predictability of their health related information is rather incorrect. This leads us to the second observation: the actual predictability of the more sensitive dimensions is higher than the perceived predictability. Vice versa, perceived predictability is higher than actual predictability for the less sensitive dimensions (with the exception of "Religious views").

It is also worth looking at any conclusions that may be reached by looking at the perceptions of individual users and in particular, users that belong to potentially sensitive groups; for instance, people that have answered that their health is poor or people that are not heterosexuals. We examined whether the sensitivity of particular dimensions differs for users belonging to different classes. We formed a two-way table with one dimension representing the class of the user (e.g. poor/good health) and the other dimension representing the sensitivity of the information. A $\mathcal{X}^2$ test was performed to examine if the perceptions of different classes of users about the sensitivity of some dimension differ. The test was positive (at the 0.05 level) for the following three dimensions: "Sexual orientation" (p-value: 0.000003), "Health factors" (p-value: 0.029) and "Religious

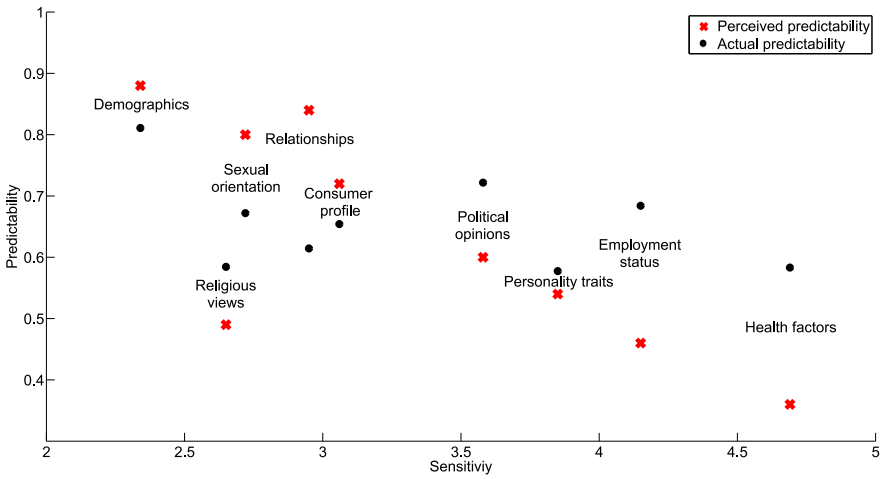

Fig. 6. Comparison of perceived and actual predictability of the disclosure dimensions with respect to sensitivity.

beliefs" (p-value: 0.011). So, for instance, homosexual and bisexual users tend to view the disclosure of information about their sexual profile as more sensitive than heterosexual users. Also, users with good health tend to view the disclosure of information about their health as less sensitive than people with poor health.

5 Conclusions

The paper discussed the issue of privacy in the context of OSNs. In particular, it examined different mechanisms by which user attributes can be predicted based on content shared by users in an OSN. Importantly, the predictability of different types of personal information was compared against the perceptions of users about the predictability and sensitivity of each type. Experiments and analysis were carried out on a dataset collected for this purpose via a custom Facebook application. The dataset consisted of the posts, images and likes of 170 Facebook users along with their responses to a survey that considered both their personal information as well as their perceptions about privacy and disclosure of information in the OSN.

A number of insights were extracted with respect to the relationship between actual predictability, perceived predictability and sensitivity. In particular, it appears that users have both correct and incorrect perceptions about the predictability of specific types of information. Moreover, the more sensitive a type of information is, the more the users underestimate its predictability. Additionally, the sensitivity of particular types of information seems to be different for users belonging to different classes. These conclusions could be useful for developing a privacy assistance tool that would support users in managing the disclosure of

personal information in online settings. For instance, assuming that a classifier predicted that a user is likely to disclose sensitive information, the user could receive an alert that his/her online sharing activities might expose unintended personal information. Recently, such a privacy assistance tool was developed in [29] in the context of photo sharing in OSNs. Extending such tools towards providing assistance for additional types of information is a promising direction for future work.

6 Acknowledgment

This work is supported by the USEMP FP7 project, partially funded by the EC under contract number 611596.

References

1. Alessandro Acquisti. The economics and behavioral economics of privacy. In Julia Lane, Victoria Stodden, Stefan Bender, and Helen Nissenbaum, editors, *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, pages 98–112. Cambridge University Press, 2014.
2. Alessandro Acquisti and Christina M Fong. An experiment in hiring discrimination via online social networks. *Available at SSRN 2031979*, 2015.
3. Lalit Agarwal, Nisheeth Shrivastava, Sharad Jaiswal, and Saurabh Panjwani. Do not embarrass: re-examining user concerns for online tracking and advertising. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, 2013.
4. Lars Backstrom and Jon Kleinberg. Romantic partnerships and the dispersion of social ties: a network analysis of relationship status on facebook. In *Proceedings of CSCW 2014*, pages 831–841. ACM, 2014.
5. David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.
6. Laura Brandimarte, Alessandro Acquisti, and George Loewenstein. Misplaced Confidences: Privacy and the Control Paradox. In *Ninth Annual Workshop on the Economics of Information Security*, page 43, Cambridge, 2010.
7. Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
8. M. D. Conover, B. Goncalves, J. Ratkiewicz, A. Flammini, and F. Menczer. Predicting the political alignment of twitter users. In *Privacy, Security, Risk and Trust (PASSAT) and SocialCom 2011*, pages 192–199, 2011.
9. Bernhard Debatin, Jennette P Lovejoy, Ann-Kathrin Horn, and Brittany N Hughes. Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences. *Journal of Computer-Mediated Communication*, 15(1):83–108, 2009.
10. Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. Liblinear: A library for large linear classification. *The Journal of Machine Learning Research*, 9:1871–1874, 2008.
11. World Economic Forum. Rethinking personal data: strengthening trust. Technical report, May 2012.
12. Yoav Freund, Robert E Schapire, et al. Experiments with a new boosting algorithm. In *ICML*, volume 96, pages 148–156, 1996.

13. Alexandru Lucian Ginsca, Adrian Popescu, Hervé Le Borgne, Nicolas Ballas, Phong Vo, and Ioannis Kanellos. Large-scale image mining with flickr groups. In *MultiMedia Modeling*, pages 318–334. Springer, 2015.
14. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.
15. Rob Heyman, Ralf De Wolf, and Jo Pierson. Evaluating social media privacy settings for personal and advertising purposes. *info*, 16(4):18–32, 2014.
16. Carter Jernigan and Behram FT Mistree. Gaydar: Facebook friendships expose sexual orientation. *First Monday*, 14(10), 2009.
17. Bart P Knijnenburg, Alfred Kobsa, and Hongxia Jin. Dimensionality of information disclosure behavior. *International Journal of Human-Computer Studies*, 71(12):1144–1162, 2013.
18. Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.
19. Michelle Madejski, Maritza Johnson, and Steven M Bellovin. A study of privacy settings errors in an online social network. In *PERCOM Workshops*, 2012.
20. H. Nissenbaum. Privacy as contextual integrity. *Wash. L. Rev.*, 79:101–139, 2004.
21. Marco Pennacchiotti and Ana-Maria Popescu. Democrats, republicans and starbucks afficionados: user classification in twitter. In *SIGKDD*, 2011.
22. Georgios Petkos, Symeon Papadopoulos, and Yiannis Kompatsiaris. Pscore: A framework for enhancing privacy awareness in online social networks. In *Availability, Reliability and Security (ARES 2015)*, pages 592–600. IEEE, 2015.
23. Sandra Sporbert Petronio. *Boundaries of privacy: dialectics of disclosure*. SUNY series in communication studies. State University of New York Press, Albany, 2002.
24. Anand S Raman, Joseph L Barloon, and Darren M Welch. Social media: Emerging fair lending issues. *The Review of Banking and Financial Services*, 28(7), 2012.
25. Delip Rao, David Yarowsky, Abhishek Shreevats, and Manaswi Gupta. Classifying latent user attributes in twitter. In *Proceedings of the 2nd international workshop on Search and mining user-generated contents*, pages 37–44. ACM, 2010.
26. Jesse Read, Bernhard Pfahringer, and Geoff Holmes. Multi-label classification using ensembles of pruned sets. In *ICDM'08*, pages 995–1000, 2008.
27. Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. Classifier chains for multi-label classification. *Machine learning*, 85(3):333–359, 2011.
28. H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9):e73791, 2013.
29. Eleftherios Spyromitros-Xioufis, Symeon Papadopoulos, Adrian Popescu, and Yiannis Kompatsiaris. Personalized privacy-aware image classification. In *Proceedings of the 6th ACM International Conference on Multimedia Retrieval*, ICMR '16, 2016.
30. Eleftherios Spyromitros-Xioufis, Grigorios Tsoumakas, William Groves, and Ioannis Vlahavas. Multi-target regression via input space expansion: treating targets as inputs. *Machine Learning*, pages 1–44, 2016.
31. Fred Stutzman, Ralph Gross, and Alessandro Acquisti. Silent Listeners: The evolution of privacy and disclosure on Facebook. *Journal of privacy and confidentiality*, 4(2):7–41, 2012.
32. Thomas Theodoridis, Symeon Papadopoulos, and Yiannis Kompatsiaris. Assessing the reliability of facebook user profiling. In *WWW*, 2015.

33. Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas. Mining multi-label data. In *Data Mining and Knowledge Discovery Handbook*, pages 667–685. Springer, 2009.
34. Alan Westin. *Privacy and freedom*. Bodley Head, London, 1970.
35. Elena Zheleva and Lise Getoor. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In *WWW*, 2009.