

Utilizing Social Networks for User Model Priming: User Attitudes

Adam Moore¹, Gudrun Wesiak², Christina M. Steiner², Claudia Hauff³,
Declan Dagger⁴, Gary Donohoe⁵ & Owen Conlan¹

¹KDEG, School of Computer Science and Statistics, Trinity College, Dublin, Ireland
{mooread,owen.conlan}@scss.tcd.ie

²Knowledge Technologies Institute, Graz University of Technology, Austria
{gudrun.wesiak,christina.steiner}@tugraz.at

³Delft University of Technology, the Netherlands
c.hauff@tudelft.nl

⁴EmpowerTheUser, Trinity Technology & Enterprise Campus, The Tower, Dublin, Ireland
declan.dagger@empowertheuser.com

⁵Department of Psychiatry, School of Medicine, Trinity College, Dublin, Ireland
DONOGHUG@tcd.ie

Abstract. Research on user modeling based on social network information has shown that some user characteristics can be accurately inferred from users' digital traces. This kind of information can be used to inform user models of adaptive systems for personalizing the system. This paper addresses a crucial question for practical application of this approach: *Are users actually willing to provide their social Web profiles and how do they perceive this?* An empirical study conducted with medical students shows that although participants are using social networks, they are reluctant about providing their identities and consider these portals rather private. The outcomes of the study uncover a clear need for further research on enhanced privacy and enhanced trust.

Keywords: user modeling, empirical study, social networks, privacy, user acceptance

1 Introduction

In our increasingly technology driven world, adaptation and personalization technologies make for a more customized, user-centric interaction with often impersonal interfaces. However, the sources of information that drives this adaptation, informing a user model that a system can utilize, are often burdensome or themselves derivative and impersonal. On the other hand, a customized experience can be created by filling in long (often deeply personal) questionnaires. In order to address this, researchers have looked to the open digital traces left on the social Web. Public and semi-public portals such as Twitter and Facebook expose personal details and preferences that can be used to inform underlying models about individuals, harvested and processed automatically and then applied to create a tailored experience.

Researchers have attempted to infer a diverse set of user characteristics, mostly from Twitter streams (due to the open nature of the portal and the ease of data collection), which have often led to algorithms with surprisingly high accuracy. In [1] the political affiliation of users in the United States was predicted with more than 90% accuracy, while in [2] the user's gender could be estimated with a similar success rate. The prediction of higher-level user characteristics, including the user's topical interests from their tweets [3] and the user's personality profile based on their Twitter [4] and Facebook [5] activities have also been investigated, though the prediction of such high-level concepts has proven to be more challenging. In all the studies presented, one important aspect of the research was the identification of the necessary user data (i.e. the user's tweets or the user's photos on Flickr) and the derivation of the ground truth (i.e. gender, political affiliation, etc.). This is usually achieved by collecting the publicly available data of random users and by manually annotating the streams with respect to the wanted characteristic(s). This means, that for such research on public streams, users are usually not explicitly asked about their willingness to participate.

This, however, leaves an open question when these user characteristics are to be employed in practice, i.e. in a working system: Are users, who use the system, actually willing to provide us with their social Web profiles? It is well known that, on the one hand, users appreciate personalized information but, on the other hand, they are very concerned about privacy and that large amounts of personal information may be tracked and made accessible to other users [6]. It has also been shown that social media are deeply integrated into users' daily lives and routines [7]. As a result, privacy attitudes (as indicated in surveys) and privacy behaviors often differ [8]. This so-called "privacy paradox" [9] is evident when comparing social network (SN) users' self-reports on their understanding of caution with regard to privacy settings and their actual lack of utilizing possibilities to change the typically very lax default settings in SNs [7]. Thus, very often the benefit of using SNs for communication or personalized contents (derived from user models) for web queries or commercial ventures outweighs the perceived privacy concerns. However, most commercial personalized web-based systems do not ask users to provide their information, but simply track them from their digital traces. Users themselves are mostly not aware of the comprehensive records search engines capture by integrating different Web 2.0 services such as Flickr, Yahoo, Twitter, etc. [10]. Thus the question arises: *"how does the privacy paradox take effect when users are explicitly asked whether their SN information may be used for personalization purposes?"*

In order to investigate students' attitudes towards providing SN information for personalizing their learning experience, we conducted a survey with medical students that were to be using an adaptive experiential training simulation, requesting a number of pieces of information on their usage and attitudes to social media.

2 Empirical Study

2.1 Method

Data on social network usage was collected in the context of a larger study on the EmpowerTheUser¹ RolePlay Simulation Platform for medical interview training.

Participants. 152 students from Trinity College Dublin participated in the study as part of their third year medical curriculum. They were sent an email requesting their participation in an online survey. 95 students (a response rate of 62.5%) filled out at least one complete section of the survey. They were on average 22.81 years old (SD = 3.79) ranging from 19 to 45 years. Half of the participants were male, half female (47 each), one participant did not indicate his or her gender.

Instruments and Procedure. Data collection was carried out over four weeks during the spring semester of 2013. Students were requested to complete the survey before starting interaction with the simulator.

Besides demographic data, a question concerning students' daily internet usage, and standard questionnaires to cover personality traits (SSP, Swedish Universities Scales of Personality [11]), learning styles (ILS, Felder-Solomon Index of Learning Styles [12]), and metacognitive awareness (MAI, Metacognitive Awareness Inventors [13]) were used. Users' perceptions of privacy, trust, and accuracy of information in Social Networks (SN) were measured by means of 12 questions. The questions differentiated between five SN: Facebook, Twitter, LinkedIn, Flickr, and MySpace.

2.2 Results

Usage. Whereas 81% of the students use Facebook, Twitter is used by only 20%, LinkedIn by 5.3% and Flickr and MySpace by only 1 person each. This basically reflects the general world wide usage of these networks². Considering the usage pattern of our sample, in the following, only results for Facebook, Twitter, and LinkedIn are reported. From the 77 Facebook users only 11 (14.3%) provided their username, Twitter and LinkedIn usernames were provided by 3 (15.8%) and one person (20%), respectively. In total IDs from 13 different persons were provided.

Figure 1(a) shows the kind of people participants intentionally interact with on different social networks. Numbers represent the percentage of account holders selecting an option. Independent of the SN used, almost all of the participants use these networks to communicate with friends (92-100%). Almost half of the Facebook users also interact with colleagues and acquaintances (all $\geq 44\%$), whereas only 26% of the Twitter users indicated to interact with those groups. On the other hand, all 5 participants with a LinkedIn account said they interact with colleagues.

Interestingly, participants who indicated using their SN accounts to interact with colleagues, acquaintances, and even everyone, were still reluctant to provide their

¹ <http://www.etu.ie/>

² <http://en.wikipedia.org/wiki/{Facebook,Twitter,LinkedIn,Flickr,MySpace}>

social network identities (SN-IDs) for research purposes. In the training scenario that followed the survey, none of the participants provided their social ID.

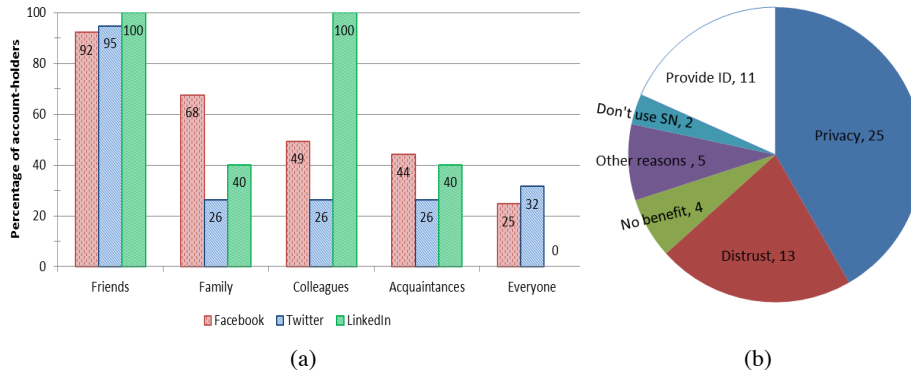


Figure 1. (a) Person-groups participants interact with in different social networks ($N = 95$) and (b) reasons for (not) providing SN-IDs (incl. frequencies of entries).

Privacy and Trust. Questions on the perception of SNs were answered by 75 to 77 students for Facebook and 18 to 29 for Twitter. In the following medians (Md) are reported as a measure for central tendency. Although students perceive their postings on social networks as rather open ($Md = 3$ on a 4-pt. scale), most of them are either 'nervous' about providing their username or simply state they would not provide it ($Md = 4$ on a 4-pt. scale). They are also rather suspicious of people and companies using their SN postings for research or commercial ventures ($Md = 4$ on a 5-pt. scale). With respect to the representation of their own personality, students think that the portrayal of their personality is partially true and that others would get a medium accurate picture of them based on their posts (both $Md = 3$ on 5-pt. scales).

Finally, students were asked about their feelings towards providing SN-IDs in order to benefit from a more personalized learning experience and whether they trust the ETU operators that their personal information will not be used for any other purpose. With $Md = 2$ (4-pt. scales) participants indicated once more that they don't feel good about providing their SN-IDs, even if it is for their own learning benefit and that they rather distrust the simulation operators. To check whether there are gender differences in the perception of SN, independent samples t-tests have been calculated. Summarizing, male participants evaluate SN-postings as less secure and private, and rate the accuracy of deducing gender as less accurate than their female colleagues.

A last question in open answer format prompted students to relate their reasoning behind how they feel about providing their SN-ID. Open responses from 60 participants were collected, of which 49 or 83% explained why they did not want to provide their SN-ID, whereas the remaining persons gave a reason for providing their ID. All answers were analyzed and sorted into 15 categories (aggregated to 6 categories in Figure 1(b)). Most comments (overall 25 entries) concerned the privacy of SN accounts, i.e. participants use them mainly to connect to friends and family (10 entries), view SNs as something private (9), and want to separate their private life from educational or business life (6). Another group of 13 students stated that they

don't know and don't trust the people behind the survey (6), that they are insecure about what happens with the information from their SN accounts (4), and that they don't want strangers going through their personal information, postings, or pictures (3). Furthermore, students commented that they don't see any benefit in using their SN information (4 entries), especially because they believe that is not related to their true or their "educational" personality, that they want to remain anonymous (2 entries), or simply that they don't see any reason to provide their ID (3). On the other hand, students who did provide their SN-IDs stated that they don't have anything to hide and that their information on the respective networks is not too personal (3 entries), or that they are simply fine with providing it (3 entries). Other users stated that they wanted to help (3) or that they hope to benefit from providing it (2). Two participants provided their Twitter or LinkedIn but not their Facebook ID (since it contains more personal information).

In order to find out how participants' attitudes are related among each other and whether there are any connections to their personality, learning style or metacognitive awareness, responses on the relevant scales were correlated by means of Spearman's Rho coefficient (for ordinal data)³. Summarized, the data show that more comfort in providing SN-ID relates to a higher perception that networks are open, more comfort with the use of information for research or commercial ventures, a better feeling of providing one's ID for a benefit regarding learning experience, and more trust that the simulator operators will not use the gained information for any other purposes. Furthermore, participants who think that SNs are very open also believe that SNs do not give a realistic, complete or accurate picture of them and have more trust in the simulator operators. Users who think that the picture derived from their posts is accurate are less comfortable with the use of their SN information for research or commercial purposes. On the other hand, participants who are comfortable with giving away information for research or commercial venture also feel good providing their ID to benefit from more personalized learning experiences and have also more trust that simulator operators will not use their information for other purposes.

A look at daily internet usage, personality traits, or learning styles did not reveal any meaningful relationships. However, students' metacognitive awareness is closely related to their trust in the simulator operators. More specifically, students who have high scores on the monitor and evaluation scales, as well as a high overall regulation of cognition score, indicated a stronger distrust in simulator operators.

Perceptions of Information Inferred from Social Network Posts. Participants were also asked to indicate how accurately they think 10 different traits can be deduced from their social network activities. Most students believe that gender, university degree course, and highest educational degree can be very accurately deduced from their SN ($Md = 5$), age, nationality, and personality somewhat accurately ($Md = 4$), whereas political convictions, income, car model, and music taste cannot be inferred from their SN ($Md \leq 2$).

³ Note: results are reported only for significant correlations derived from Spearman's Rho (with $p < .05$); correlations are either for Facebook, Twitter, or both.

3 Discussion and Conclusion

From a sample of nearly 100 respondents, it became clear that, although they are active on social networks, they do not consider them a place for information to be gathered that could be useful in tailoring training to their individual needs.

With metacognitive awareness being positively correlated with a definitive unwillingness to share this information, there is clearly a need to find ways to increase the trust learners have in the people behind the learning environment they are using.

It is interesting to note the perception of both the privacy and information that can be derived from an active social network account. For the majority of traits, the participants' intuition about how well they can be estimated from the SN is in line with existing research and SN are perceived as rather open. Nevertheless, privacy is a great concern. This, therefore, presents somewhat of a dilemma for researchers and practitioners in adaptation technologies. We can now, with reasonable accuracy, infer and predict many aspects of our systems' users from their traces on open, publically available channels. However, when directly questioned about this approach, our users are reluctant to disclose their identities within these networks (information that can often actually be obtained without their consent), express discomfort and, when asked directly in the training simulator to provide this information for an illustrated educational benefit, exactly zero of our cohort of 152 did so. Thus, in contrast to the privacy paradox concerning users' reported attitudes and behavior [8][9], our sample was very consistent in their reported unwillingness to provide their SN-information and their actual behavior. The paradox, though, lies in the fact that although users are willing to disclose personal information on their SN, they feel uncomfortable providing this information to personalize their learning. Clearly, more work is needed to bridge the gap between perceived usage and audience of these portals and those hoping to use the information contained within to provide benefit to its participants.

In summary, it seems that our study participants view their SN mainly as a means to connect with friends and family and thus as something that should not be linked to their professional development and training. In the same line, willingness to provide their SN-IDs is closely related to the belief, that the networks are very open anyway, and that their true personality cannot be derived from their posts. Personality typing gave entirely normative responses, with no indication of overly cautious or private characteristics. However, it is known that privacy concerns increase with higher age, education, and income [6]. The reluctance about utilizing SN information for user modeling might, to some extent also be correlated with the students' background; thus investigating cohorts from different disciplines, like computer or information science, would be desirable. In line with [6], in order to use information from SN portals, it seems key to explicitly explain to users what kind of information is used, how the information is extracted, and how exactly participants could benefit from providing their network IDs. Also knowledge and control over the used information fosters users' willingness to disclose personal information. In addition some basic information about those making use of the information should help to build up trust into the diligent handling of their information. Future research needs to focus on conclusive ways to convey the benefits for the users and to give them more control and insight on the actually utilized body of information.

Acknowledgments. The research leading to these results has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement no 257831 (ImREAL project).

References

1. Conover, M. D., Gonçalves, B., Ratkiewicz, J., Flammini, A., Menczer, F.: Predicting the political alignment of users. In *SocialCom/PASSAT*, pp. 192-199. IEEE (2011)
2. Burger, J. D., Henderson, J., Kim, G., Zarrella, G.: Discriminating gender on Twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1301-1309. Association for Computational Linguistics (2011)
3. Michelson, M., Macskassy, S. A.: Discovering users' topics of interest on Twitter: a first look. *Proc 4th w/s on Analytics for noisy unstructured text data*, pp. 73-80. ACM (2010)
4. Quercia, D., Kosinski, M., Stillwell, D., Crowcroft, J.: Our Twitter profiles, our selves: Predicting personality with Twitter. In *SocialCom/PASSAT*, pp. 180-185. IEEE (2011)
5. Wald, R., Khoshgoftar, T., Sumner, C.: Machine prediction of personality from Facebook profiles. In *13th International Conference on Information Reuse and Integration (IRI)*, pp. 109-115. IEEE (2012)
6. Kobsa, A.: Privacy-Enhanced Web Personalization. In P. Brusilovsky, A. Kobsa, W. Nejdl (eds.). *The Adaptive Web: Methods and Strategies of Web Personalization*. Berlin, Heidelberg, New York: Springer Verlag, pp. 628-670 (2007).
7. Debatin, B., Lovejoy, J.P., Horn, A.-K., Hughes, B.N.: Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences. *Journal of Computer-Mediated Communication* 15, pp 83-108 (2009).
8. Stutzman, F., Kramer-Duffield, J.: Friends Only: Examining a Privacy-Enhancing Behavior in Facebook. *CHI*, pp. 1553-1562 (2010)
9. Barnes, S.B.: A privacy paradox: Social networking in the United States. *First Monday* 11 (2006). <http://firstmonday.org/ojs/index.php/fm/article/view/1394/1312#b1>.
10. Zimmer, M.: The Externalities of Search 2.0: the Emergin Privacy Threats when the Drive for the Perfect Search Engine meets Web 2.0. *First Monday* 13 (2008). <http://firstmonday.org/ojs/index.php/fm/article/view/2136/1944>
11. Gustavsson, J.P., Bergman H., Edman, G., Ekselius, L., von Knorring, L., Linder, J.: Swedish universities Scales of Personality (SSP): construction, internal consistency and normative data. *Acta Psychiatrica Scandinavica* 102, 217-225 (2000)
12. Felder, R.M., Spurlin, J.: Applications, Reliability and Validity of the Index of Learning Styles. *International Journal of Engineering Education* 21, 103-112 (2005)
13. Schraw, G., Dennison, R.S.: Assessing metacognitive awareness. *Contemporary Educational Psychology* 19, 460-475 (1994)