Improving Social Recommendations Through Privacy-Awareness

Alexander Korth and Andreas Nürnberger
Otto-von-Guericke Universität Magdeburg, Data & Knowledge Engineering Group
Magdeburg, Germany
{alexander.korth,andreas.nuernberger}@ovgu.de

ABSTRACT
Contemporary Social Networking Services lack proper transparency and control features for users to understand and to efficiently steer the reach of their content and profile information, i.e. their online privacies. In this paper, we outline the problems and explain different access control mechanisms to approach this field. We briefly introduce our approach of a platform that provides privacy awareness by design and list a selection of the applied features for transparency and control. We report of a quantitative user survey which proved our platform to be understandable by users. Compared to conventional social networking services, users found it easier to understand both, whom is granted access to information, and how to control this. We show that through the provided and improved transparency and control our users sensed a) other users to be closer friends and b) content and social recommendations to be more relevant.

Author Keywords
privacy, control, transparency, recommendation, social networking service

ACM Classification Keywords
H.4.3 Information Systems Applications: Communications Applications

INTRODUCTION

Although there is no single definition or meaning of the term *privacy* (from Latin: *privatus*) [15], individual privacy is commonly understood as the ability of individuals to selectively reveal information about themselves to others. In his book Privacy and Freedom [16], Westin defines:

> Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others.
> (Alan Westin)

Workshop SRS11, March 19-23, 2011, Hangzhou, China. In Conjunction with CSCW 2011.

In the realm of Web applications, and especially Social Networking Services (SNS), this definition makes clear that it is the provided privacy features for personal information and content that decide over how well users perceive their privacies to be protected at a given service. Accordingly, to protect their privacies online, besides a level of trust users need features to overlook and steer the reach of their information. Therefore, they need transparency and control, respectively.

In the offline world, humans have senses for transparency, and contexts on which they apply their means of control over audience, and content of their communication on. Prior works have defined two context areas humans use to manipulate communication. Firstly, Communication Context is a set of attributes that affect the content of communication between two individuals [9, 5]. The attributes include the history and effect of information and receivers on the actual content that is about to be communicated. Secondly, Social Context which refers to who is communicating with whom and the influence of the strength of relationship shared between them [4]. To master these, humans have the needed sensors and knowledge of their cultural and social norms for communication to preserve the privacies of themselves and others while socializing.

For instance, imagine someone who wants to tell a vis-à-vis friend about a secret beach in a spoken dialog. The sender can see nearby people and adjust its volume of speech. From its common sense knowledge it can judge if the dialog may be recorded and grant access to unwanted, further audience later. It can evaluate if eventually close-standing, foreign listeners may cause a threat to their privacies if they receive the communication. On this basis of a sensed and controllable Social Context, Communication Context is applied to adjust the content. However, if the sender's decisions prove wrong its own or the privacy of receivers or people named in the dialog may be threatened or hurt. The basic source for privacy-related concerns and problems in the online world is a lack of sensors and common sense knowledge about the norms and laws of online communication. These norms and laws are determined by every SNS's provider's access control policies to user-generated content.

More generally, the privacy-related problems found in SNSs base on a conflict of interests between users and providers. As users are driven by their need to socialize online [7, 10] and therefore need transparency and control mechanisms to protect their privacies, commercial providers quest to in-

crease their profit [12]. Therefore, providers maximize information flow to grow their user- and content-bases. Tools provided to users for transparency and control of personal information flow are accordingly intentionally not present, not promoted or unusable. If no sufficient transparency is given for where information flows, users can neither learn nor control whom is granted access to their content. As users naturally try to adopt social norms for communicating from the offline world to the online counterpart, the lack of such tools is a common source for misunderstandings and threats to the privacy of posting users and those affected.

Our approach aims at providing a tool that is easier to understand and use by users to enable them to communicate online at a greater certainty of privacy protection. However, resolving the named conflict of interest between users and providers is out of scope of this work.

Our hypothesis is if users are provided with features for transparency and control in the online world (preconditioned a necessary level of trust in the service [2]), they are not only intuitively able to protect their privacies more efficiently but also content relevancy as perceived by receivers consequently increases as a by-product which we understand as social recommendation.

ACCESS CONTROL LISTS

Access Control Lists (ACL) are the provider's basic tool for implementing its rules for who is granted access to what information. Hence, we briefly want to give an introduction into the topic.

Access Control Policies (ACP) define what kind of access is granted to which data to whom under what conditions. The content in question to be accessed can be a file, a database record, a Website, also a SNS profile page, or a service to be called. Pekárek and Pötzsch [13] provide a general list of different access control mechanisms for use in SNSs and collaborative workspaces (CW):

Access based on identifier: The access to a piece of information is granted based on the requestor's identifier, i.e. their user ID or handle. E.g., a user's profile page or personal content is accessible for befriended users in a SNS.

Access based on roles: Many CWs and some SNSs assign roles to their users to, e.g., grant group founders additional rights compared to regular members.

Access based on groups: Groups are a set of user IDs and thus a proper tool of granting a collection of users access to information.

Access based on properties: This access mechanism uses properties of the accessing subject, e.g. its age, or the object about to be accessed, to decide over granting access.

Access based on context: This mechanisms uses the context of the access, e.g. the time or location but also system state like its load, to decide over the access being granted.

ACPs are the foundation for ACLs which combine and weight ACPs to a defined rule or list of rules to define whom is granted access to a particular piece of information. SNS providers use ACLs to implement their platform's rules for information flow. E.g., Perez reported of Facebook's ACL granting users access to photos [14].

However, the selection, implementation and presentation of ACLs which providers realize on their SNSs decide over how intuitive a service is regarding online communication and how easy to learn its rules are. Most SNS providers use *access based on identifier* for most of their access decisions. This, on the one hand, makes a service easy to understand because basically all content is accessible to all befriended users, but, on the other hand, does not provide a sufficient means for a more selective online communication.

Seemingly, users do not need to fully understand the mechanisms of a service but they need a basic feeling of trust in the provider [2]. As most users do neither read nor understand privacy policies [1, 11], most of them have their privacy concerns satisfied by providers advertising privacy features, no matter if they use them or if they are usable at all [3]. Open defaults for information flow controls and unusable privacy controls ensure maximized information sharing among the majority of users which is good for the growth of the provider's service. This is the unsatisfactory way the named conflict of interests between users and providers is usually handled as of today.

A PRIVACY-BY-DESIGN PLATFORM APPROACH

Having said all that, our SNS approach faces the challenge of, on the one hand, providing privacy awareness by transparency and control features, and, on the other hand, remaining intuitive and easy to understand and learn to users. Therefore, we chose to implement *access based on groups* as a basic ACP since groups are a social concept people know from the offline world and learned how to deal with regarding disclosing information while preserving privacies.

Our platforms consists of three primary entities: Groups, users and tips. Tips are an exemplary content entity which users create, share with and recommend to friends. Our groups consist of members, i.e. users, and content, i.e. tips. People use these groups as a means to *control* who is granted access to their content by a) assigning content to groups making it accessible to all current and future members, and b) by inviting further users to that groups. Group administrators additionally have the right to de-assign users and content from groups.

By this ACP and in contrast to the commonly used *access based on identifier*, people are furthermore empowered to maintain different personae by controlling their Communication Contexts throughout heterogenous groups. E.g., users can play their business persona in one group by sharing job-related content with colleagues whilst freely posting private, more intimate, topics to friends and family in another. As said before, this is difficult in SNSs that implemented *access based on identifier*.

Consequently, our system does not provide any groups that are open to join. All groups are closed and joinable by invitation only. Thus, users have a Social Context to base a decision on what to publish to that group, affected by its members. Furthermore, the relatively static Social Context lessens concerns about who will have access to this information in the future. Furthermore, group administrators have the option to forbid normal group members to invite further users.

To conclude, users are provided with groups, including all needed functionality to assign and de-assign content and to manage their own memberships, as a means to *control* the reach and information flow of their content.

To provide the needed *transparency*, we implemented a number of features which mainly play a role in the area of human-computer interaction.

First of all, we implemented omnipresent lists of members of groups. Like this, users can oversee to whom their content is made accessible. This transparency is of great importance since in the online world trust in online interaction has been shown to be of lesser felt necessity than in face to face encounters [6]. Without such transparency, people tend to forget which users they are about to make content accessible to.

To lessen the learning effort to use our platform, we iteratively improved its usability by qualitative user interviews. E.g., we developed an Information Architecture and Navigation Design that allows users to easily access all members and tips that are accessible through their group memberships instead of forcing them to navigate through groups in order to reach a user or tip. Being allowed to browse users and content on a meta-group level eases the navigation to information and allows for sophisticated filter, sort and search operations. Further examples for improving the usability and intuitiveness of our platform include the *typing* of entities by, e.g. color-coding and assigning particular avatar image ratios: Portrait for users, square for groups, and landscape for tips.

EVALUATION RESULTS

A user survey consisting of 20 questions was performed amongst the users of our SNS platform ($N = 67$). Although the sample is rather small for a quantitative analysis, it allows tendency conclusions.

The majority of respondents (ca. 61%) were classified as Privacy Pragmatists [8]. Most of them are male (87%) and between 26 and 35 years of age (71%). They earn an average of ca. 47,000 Euros annually and most possess an academic degree (75%). 70% actively use the Internet between 2–8 hours and 43% use SNSs actively for about 0.5–1 hours every day.

One block of survey questions (SQ) researched the understanding of what the platform is good for, how information is organized, and how to navigate through information. The vast majority of users (77%) found it easy to understand what to do with the platform (SQ 4). 63% of respondents answered that they have understood the applied Information Architecture, i.e. the way that application entities are related to each other (SQ 5). 69% of participants verified the intuitiveness of how to navigate through information, i.e. the Navigation Design (SQ 6). The three questions show significant correlations (Table 1). The numbers show that the majority of users understood the platform's purpose and how to use it. On the other hand, it is obvious that there is room for improvements to make the application understandable to more users.

		SQ 4	SQ 5	SQ 6
SQ 4	Corr. coeff.	1.000	.432**	.362**
	Significance		.000	.001
	N	67	67	67
SQ 5	Corr. coeff.	.432**	1.000	.486**
	Significance	.000		.000
	N	67	67	67
SQ 6	Corr. coeff.	.362**	.486**	1.000
	Significance	.001	.000	
	N	67	67	67

Table 1. Correlations (Kendall's τ) between questions 4–6. The same group of users understood the platform's value proposition, its Information Architecture and Navigation Design. Two asterisks (**) mark correlations significant at the 0.01 level (2-tailed).

The next block of questions interviewed the participants regarding the provided transparency and control features. Survey Question 8 resulted in a 60% of users who approved that they understood the provided features to *control* the audience of their content. Survey Questions 9 and 10 interviewed the respondents concerning the *transparency* features. 48% agreed on a provided transparency of who has access to published information (SQ 9). A similar result (51%) was observed asking users if they feel that they can communicate openly to their friends because they have the transparency of who has access (SQ 10).

Table 2 lists correlations between questions 4–6 and questions 8–10. The numerous significant correlations prove that users who understood the platform's value proposition, Information Architecture, and Navigation Design also understood its control features and especially its transparency features. Consequently, it must be assumed that improving the intuitiveness of the platform will result in a higher understanding of control and transparency features.

The evaluation so far showed that about half of users were effectively empowered with transparency and control features to overview and steer the audience of published information, i.e. their privacy.

In the following, our platform is compared to other SNSs, namely Facebook and XING. Facebook is used by 55 of the respondents, XING by 46[1]. Concerning the understandabil-

[1]The survey included further SNSs, i.e. StudiVZ/MeinVZ, Lokalisten, Wer Kennt Wen, LinkedIn and MySpace, but these platforms were not sufficiently used by the survey participants.

		SQ 8	SQ 9	SQ 10
SQ 4	Corr. coeff.	.169	.273*	.440**
	Significance	.132	.015	.000
	N	62	61	57
SQ 5	Corr. coeff.	.153	.308**	.395**
	Significance	.166	.005	.001
	N	62	61	57
SQ 6	Corr. coeff.	.228*	.384**	.361**
	Significance	.039	.001	.002
	N	62	61	57

Table 2. Correlations between questions 4–6 related and questions 8–10. Users who understood the platform's value proposition, Information Architecture, and Navigation Design also understood its control and especially its transparency features. One asterisk (*) marks correlations significant at the 0.05 level (2-tailed), two asterisks mark those significant at the 0.01 level (2-tailed).

ity of provided control features (SQ 8), our platform scores about the same compared to Facebook and XING. Similar results were observed through Survey Questions 9 and 10 which interrogated users about the perceived transparency of information reach when posting content online (Table 3).

	SQ 8	SQ 9	SQ 10
Platform	59.68%	47.54%	50.88%
Facebook	65.63%	42.86%	52.54%
XING	63.46%	49.02%	49.94%

Table 3. Seemingly, the respondents scored (numbers indicate percentage of users that agreed on the features' presence) the perceived features for control (SQ 8) and transparency (SQs 9 and 10) almost equally between our approach, Facebook and XING.

Interestingly, the respondents rating our platform's control (Table 4) and transparency (Tables 5 and 6) *high* did not answer so for Facebook's and XING's. The latter two SNSs' control and transparency features, on the other hand, show a high correlation regarding respondents rating those features.

Survey Question 8		Platform	Facebook	XING
Platform	Corr. coeff.	1.000	.076	.040
	Significance		.499	.753
	N	62	60	48
Facebook	Corr. coeff.	.076	1.000	.386**
	Significance	.499		.001
	N	60	64	51
XING	Corr. coeff.	.040	.386**	1.000
	Significance	.753	.001	
	N	48	51	52

Table 4. Correlations between our platform, Facebook and XING in Survey Question 8. While there is no correlation between our platform and the other two, Facebook and XING correlate significantly.

Before was shown that the fraction of respondents that understood our platform as such (SQs 4–6), also rated its control and transparency features as being effective. Since there is no correlation between respondents that rated high our platform and Facebook (which correlates significantly with XING) it can be concluded that the users that rated our platform high regarding its control and transparency features, rated Facebook's and XING's low or were unsure. This indicates that our platform features for user privacy preservation are rated higher than those of Facebook and XING by respondents.

Survey Question 9		Platform	Facebook	XING
Platform	Corr. coeff.	1.000	.160	-.077
	Significance		.153	.538
	N	61	58	48
Facebook	Corr. coeff.	.160	1.000	.355**
	Significance	.154		.002
	N	58	63	50
XING	Corr. coeff.	-.077	.355**	1.000
	Significance	.538	.002	
	N	48	50	51

Table 5. Correlations between our platform, Facebook and XING in Survey Question 9. There is no correlation between our platform and the other two; Facebook and XING correlate significantly.

Survey Question 10		Platform	Facebook	XING
Platform	Corr. coeff.	1.000	-.030	-.090
	Significance		.796	.496
	N	57	54	44
Facebook	Corr. coeff.	-.030	1.000	.421**
	Significance	.796		.001
	N	54	59	46
XING	Corr. coeff.	-.090	.421**	1.000
	Significance	.496	.001	
	N	44	46	47

Table 6. Correlations between our platform, Facebook and XING in Survey Question 10. No correlation could be found between our platform and the other two, but Facebook and XING correlate significantly.

Increased Content Relevancy

Through Survey Question 10, more than half of the respondents (51%) stated that through the given transparency of the audience for their information and content they feel empowered to communicate more openly to their friends. In Survey Question 11, two-thirds of users (66%) answered their friends on our platform were mainly real friends. Survey Question 12 reports that 72% of users know the majority of group members on our platform in person. These results indicate that users perceive an improved Social Context on our platform.

Survey Question 13 proves that the content shared throughout the groups of our platform is, based on the transparency and control features, highly relevant to users because it comes from real friends witnessing an increased Communication Context. The respondents that *agreed* on Survey Question 13 are the same that understood the platform's purpose and functioning. Table 7 lists the significant correlations between the respondents of the relevant questions. As said before, it must be assumed that the fraction of users who agreed on SQ 13 can be increased by improving the understandability of the platform.

The comparison with other SNSs, i.e. Facebook and XING, has analogies to the analysis of the beforehand section (Table 8). Whereas Facebook and XING correlate significantly

		SQ 4	SQ 5	SQ 6
SQ 13	Corr. coeff.	.293*	.376**	.251*
	Significance	.010	.001	.024
	N	56	56	56

Table 7. Significant correlations between questions 4–6 related and question 13 state that users who understood our platform's value proposition, Information Architecture, and Navigation Design also found its content relevant since it comes from real friends.

positive, Facebook and our platform do not correlate. Interestingly, XING and our platform correlate significantly negative. This indicates that users who sense a higher relevancy of content on our platform answered that they do not feel so on XING.

Survey Question 13		Platform	Facebook	XING
Platform	Corr. coeff.	1.000	-.035	-.292*
	Significance		.764	.032
	N	56	50	38
Facebook	Corr. coeff.	-.035	1.000	.296*
	Significance	.764		.026
	N	50	55	40
XING	Corr. coeff.	-.292*	.296*	1.000
	Significance	.032	.026	
	N	38	40	42

Table 8. Correlations between our platform, Facebook and XING in Survey Question 13. While Facebook and XING correlate significantly positive, Facebook and our platform do not correlate. Interestingly, XING and our platform correlate significantly negative.

CONCLUSION

In this paper, we have outlined the problems of users protecting their privacies online. We have sketched the state of the art of problems of contemporary SNSs and pointed at functional lacks for users to oversee and control the reach of their published content and personal information.

Consecutively, we have introduced different Access Control Policies, providers can base their Access Control Lists on, to motivate our own approach which bases access permissions on groups rather than friendship links. Whereas both are an intuitively understandable means for a realization, we believe that groups allow for a better control of Social Contexts for, e.g., different types of content. Consequently, we have chosen closed groups as the Access Control Policy for our own approach because it promises usable Social Contexts for an improved Communication Context.

We have shown that by the provision of a greater privacy awareness through transparency and control features a higher content relevancy is perceived by users. Although we detected room for improvements, the significance of correlations clearly showed the connections of the two areas privacy and content quality.

REFERENCES

1. A. I. Antón, E. Bertino, N. Li, and T. Yu. A roadmap for comprehensive online privacy policy management. *Communications of the ACM*, 50(7):109–116, 2007.
2. G. Bansal, F. Zahedi, and D. Gefen. The moderating influence of privacy concern on the efficacy of privacy assurance mechanisms fo building trust: A multiple context investigation. In *Proc. ICIS*, 2008.
3. J. Bonneau and S. Preibusch. The privacy jungle: On the market for data protection in social networks. In *Proc. WEIS*, 2009.
4. M. D. Choudhury, H. Sundaram, A. John, and D. Seligmann. Dynamic prediction of communication flow using social context. In *Proc. HT*, 2008.
5. M. D. Choudhury, H. Sundaram, A. John, and D. D. Seligmann. Contextual prediction of communication flow in social networks. In *Proc. WI*, 2007.
6. C. Dwyer, S. R. Hiltz, and K. Passerini. Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. In *Proc. AMCIS*, 2007.
7. H. Krasnova, T. Hildebrand, O. Günther, A. Kovrigin, and A. Nowobilska. Why participate in an online social network: An empirical analysis. In *Proc. ECIS*, 2008.
8. P. Kumaraguru and L. F. Cranor. Privacy indexes: A survey of Westin's studies. Technical report, Institute for Software Research International, School of Computer Science, December 2005.
9. A. Mani and H. Sundaram. Modeling user context with applications to media retrieval. *Multimedia Systems*, 12(4-5):339–353, 2007.
10. A. H. Maslow. A theory of human motivation. *Psychological Review*, 50(4):370–396, 1943.
11. G. Milne and M. Culnan. The culnan milne survey on consumers online privacy notices. *Journal of Interactive Marketing*, 18(3), 2004.
12. T. O'Reilly. What is Web 2.0: Design patterns and business models for the next generation of software. `http://oreilly.com/web2/archive/what-is-web-20.html`, September 2005.
13. M. Pekárek and S. Pötzsch. Requirements and concepts for privacy-enhancing access control in social networks and collaborative workspaces. `http://www.primelife.eu/images/stories/deliverables/h1.2.5-requirements_selective_access_control-public.pdf`, July 2009.
14. S. Perez. 8 steps to facebook photo privacy, according to facebook engineer (we're still confused). `http://www.readwriteweb.com/archives/8_steps_to_facebook_photo_privacy_according_to_facebook_engineer.php`, December 2010.
15. The Stanford Encyclopedia of Philosophy. Privacy. `http://plato.stanford.edu/entries/privacy`, September 2006.
16. A. F. Westin. *Privacy and Freedom*. The Bodley Head, 1967.