DSIC
DEPARTAMENTO DE SISTEMAS
INFORMÁTICOS Y COMPUTACIÓN

UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

GTI IA
Grupo de Tecnología Informática
Inteligencia Artificial

PhD THESIS

Enhancing Privacy Management on Social Network Services

Author: Ricard López Fogués

Advisors: Dr. Jose M. Such Aparicio
Dr. Agustín Espinosa Minguet
Dra. Ana García-Fornes

Departamento de Sistemas Informáticos y Computación,
Universidad Politècnica de València,
Valencia, Spain

March 2017

A tota la meua família;
als més majors (Lucia, mamà, José Luis i Francesc)
i a Jr, que, fins d'ací a uns anys,
no podrà llegir aquesta dedicatòria.
Sense vosaltres, cap esforç tindria sentit.

Acknowledgments

En primer lloc, voldria agrair la inestimable ajuda dels meus directors de tesi Jose, Ana i Agustín. M'han servit de guia i suport durant els meus estudis doctorals i han permés que no perdera el rumb en el meu treball d'investigació. Durant aquests anys m'han demostrat una gran qualitat humana; sempre han estat ahí quan els he necessitat i sempre disposats a tirar una mà. Sense ells, aquesta tesi no existiria.

I would also like to thank Munindar Singh and Pradeep Murukannaiah for their guidance and help during my stay at NC State. They were great hosts during my two years in Raleigh and two of the best researchers I have ever worked with. I hope that we can keep working together for years to come.

Durant els meus anys com a estudiant de doctorat he conegut a gent excepcional als que considere molt més que companys de laboratori. Agrair a tota la gent dels laboratoris 108 i 205 (Bexy, Raúl, Mario, Joan Andreu, Joan Bellver, Juanmi, Ana, Vero i Matthijs) els meravellosos anys que vaig viure al DSIC. Mai oblidaré els viatges del GTI, els dinars a la vella, les cerveses, els sopars i tot el que hem viscut dins i fora del treball.

También quiero agrader a los Ralencianos, parte fundamental de mi aventura en USA, conseguir que mis años allí fueran inolvidables. Rita, Manolo, Adri, Lola, Iván, Kostas, Àngels, y Kenneth; imposible olvidar los interminables cafés en la facultad, los partidos de volley playa (with chips), las barbacoas en Kensington Park, los domingos de Juego de Tronos, y tantas otras cosas que aquí no caben.

Finalment, agrair la confiança que el grup GTI-IA ha dipositat en mi. Tots els anys que vaig passar com a becari al DSIC, vaig sentir que formava part d'una gran família. Un plaer haver treballat amb tots vosaltres.

Resumen

En los últimos años, los servicios de redes sociales, como Facebook o LinkedIn, han experimentado un crecimiento exponencial. Los usuarios valoran positivamente sus muchas funcionalidades tales como compartir fotos, o búsqueda de amigos y trabajo. En general, los usuarios aprecian los beneficios que las redes sociales les aportan. Sin embargo, mientras el uso de redes sociales se ha convertido en rutina para mucha gente, brechas de privacidad que pueden ocurrir en redes sociales han aumentado los recelos de los usuarios. Por ejemplo, es sencillo encontrar en las noticias casos sobre personas que han perdido su empleo debido a algo que compartieron en una red social. Para facilitar la definición de los ajustes de privacidad, los proveedores de servicios emplean controles de acceso sencillos que normalmente se basan, de forma exclusiva, en listas o círculos de amigos. Aunque estos controles de acceso son fáciles de configurar por un usuario medio, investigaciones recientes indican que éstos carecen de elementos tales como la intensidad de los vínculos personales, que juegan un papel clave en cómo los usuarios deciden qué compartir y con quién. Además, a pesar de la simplicidad de los controles de acceso, investigaciones sobre privacidad en redes sociales señalan que los usuarios han de esforzarse para controlar de forma efectiva como su información fluye en estos servicios.

Para ofrecer a los usuarios un marco de privacidad más robusto, trabajos recientes proponen un nuevo paradigma para controles de acceso basado en relaciones. A diferencia de los controles de acceso tradicionales donde los permisos se otorgan en base a usuarios y sus roles, este paradigma emplea elementos sociales como la relación entre el propietario de la información y su audiencia potencial (por ejemplo, sólo mis hermanos pueden ver la foto). Los controles de acceso que siguen este paradigma ofrecen a los usuarios mecanismos para el control de la privacidad que representan de una forma más natural como los humanos razonan sobre cuestiones de privacidad. Además, estos controles de acceso pueden lidiar con problemáticas específicas que presentan las redes sociales. Específicamente, los usuarios comparten de forma habitual información que atañe a muchas personas, especialmente a otros miembros de la red social. En tales situaciones, dos o más personas pueden tener preferencias de privacidad que entran en conflicto. Cuando esto ocurre, no hay una configuración correcta de privacidad que sea evidente. Estas situaciones son normalmente identificadas como escenarios de privacidad multiusuario.

Dado que los controles de acceso basados en relaciones son complejos para el usuario promedio de redes sociales, los proveedores de servicios no los han adoptado. Por lo tanto, para permitir la implementación de tales controles de acceso en redes sociales actuales, es necesario que se ofrezcan herramientas y mecanismos que faciliten su uso. En este sentido, esta tesis presenta cinco contribuciones: (1) una revisión del estado del arte en manejo de privacidad en redes sociales que permite identificar los retos más importantes en el campo, (2) BFF, una herramienta para obtener automáticamente la intensidad de los vínculos personales y las comunidades de usuarios, (3) un nuevo control de acceso que emplea comunidades, identificadores individuales, la intensidad de los vínculos personales, y etiquetas de contenido, (4) un modelo novedoso para representar y razonar sobre escenarios de privacidad multiusario que emplea tres tipos de características: factores contextuales, preferencias de usuario, y argumentos de usuario; y, (5) Muppet, una herramienta que recomienda configuraciones de privacidad en escenarios de privacidad multiusuario. Las contribuciones de esta tesis emplean técnicas de inteligencia artificial tales como aprendizaje automático, minería de datos, y colaboración distribuida. Las contribuciones han sido validadas por medio de estudios con participantes humanos. Concretamente, tres estudios con 38, 50, y 988 participantes han aportado los datos que han sido empleados para evaluar cada contribución. Los resultados muestran que los controles de acceso para redes sociales pueden ser mejorados mediante herramientas que automatizan tareas relacionadas con privacidad (BFF y Muppet) y modelos formales que representan fidedignamente cómo los humanos razonan sobre la privacidad.

Resum

En els darrers anys, els servicis de xarxes socials, com Facebook o LinkedIn, han experimentat un creixement exponencial. Els usuaris valoren positivament les seues variades funcionalitats com la compartició de fotos o la cerca d'amics i treball. En general, els usuaris aprecien els beneficis que les xarxes socials els aporten. No obstant això, mentre l'ús de les xarxes socials s'ha convertit en rutina per a molta gent, bretxes de privacitat que poden ocórrer en xarxes socials han augmentat els recels dels usuaris. Per exemple, és senzill trobar notícies sobre persones que han perdut el seu treball per alguna cosa que compartiren a una xarxa social. Per facilitar la definició dels ajustos de privacitat, els proveïdors de servicis empren controls d'accés senzills que normalment es basen, de forma exclusiva, en llistes o cercles d'amics. Encara que aquests controls d'accés són fàcils d'emprar per a un usuari mitjà, investigacions recents indiquen que aquests manquen elements com la força dels vincles personals, que juguen un paper clau en com els usuaris decideixen què compartir i amb qui. A més a més, malgrat la simplicitat dels controls d'accés, investigacions sobre privacitat en xarxes socials revelen que els usuaris han d'esforçar-se per a controlar de forma efectiva com fluix la seua informació en aquests servicis.

Per a oferir als usuaris un marc de privacitat més robust, treballs recents proposen un nou paradigma per a controls d'accés basat en relacions. A diferència dels controls d'accés tradicionals on els permisos s'atorguen segons usuaris i els seus rols, aquest paradigma empra elements socials com la relació entre el propietari de la informació i la seua audiència potencial (per exemple, sols els meus germans poden veure aquesta foto). Els controls d'accés que segueixen aquest paradigma ofereixen als usuaris mecanismes per al control de la privacitat que representen d'una forma més natural com els humans raonen sobre la privacitat. A més a més, aquests controls d'accés poden resoldre problemàtiques específiques que presenten les xarxes socials. Específicament, els usuaris comparteixen de forma habitual informació que concerneix moltes persones, especialment a altres membres de la xarxa social. En aquestes situacions, dues o més persones poden tindre preferències de privacitat que entren en conflicte. Quan açò ocorre, no hi ha una configuració de privacitat correcta que siga evident. Aquestes situacions són normalment identificades com escenaris de privacitat multiusari.

Donat que els controls d'accés basats en relacions són complexos per a l'usuari mitjà de xarxes socials, els proveïdors de servicis no els han adoptat. Per tant, per a permetre la implementació d'aquests controls d'accés en xarxes socials actuals, és necessari oferir ferramentes i mecanismes que faciliten el seu ús. En aquest sentit, aquesta tesi presenta cinc contribucions: (1) una revisió de l'estat de l'art en maneig de privacitat en xarxes socials que permet identificar els reptes més importants en el camp, (2) BFF, una ferramenta per a obtenir automàticament la força dels vincles personals i les comunitats d'usuaris (3) un nou control d'accés que empra comunitats, identificadors individuals, força dels vincles personals, i etiquetes de contingut, (4) un model nou per a representar i raonar sobre escenaris de privacitat multiusari que empra tres tipus de característiques: factors contextuals, preferències d'usuari, i arguments d'usuaris; i, (5) Muppet, una ferramenta que recomana configuracions de privacitat en escenaris de privacitat multiusuari.

Les contribucions d'aquesta tesi empren tècniques que inclouen intel·ligència artificial, aprenentatge de màquines, mineria de dades, i col·laboració distribuïda. Les contribucions han sigut validades mitjançant estudis amb participants humans. Concretament, tres estudis amb 38, 50, i 988 participants han aportat les dades que han sigut utilitzades per a avaluar cada contribució. Els resultats mostren que els controls d'accés per a xarxes socials poden ser millorats mitjançant ferramentes que automatitzen tasques relacionades amb privacitat (BBF i Muppet) i models formals que representen fidedignament com els humans raonen sobre la privacitat.

Summary

In the recent years, social network services, such as Facebook or LinkedIn, have experienced an exponential growth. People enjoy their functionalities, such as sharing photos, finding friends, looking for jobs, and in general, they appreciate the social benefits that social networks provide. However, as using social network has become routine for many people, privacy breaches that may occur in social network services have increased users' concerns. For example, it is easy to find news about people being fired because of something they shared on a social network. To enable people define their privacy settings, service providers employ simple access controls which usually rely exclusively on lists or circles of friends. Although these access controls are easy to configure by average users, research literature points out that they are lacking elements, such as tie strength, that play a key role when users decide what to share and with whom. Additionally, despite the simplicity of current access controls, research on privacy on social media reports that people still struggle to effectively control how their information flows on these services.

To provide users with a more robust privacy framework, related literature proposes a new paradigm for access controls based on relationships. In contrast to traditional access controls where permissions are granted based on users and their roles, this paradigm employs social elements such as the relationship between the information owner and potential viewers (e.g., only my siblings can see this photo). Access controls that follow this paradigm provide users with mechanisms for disclosure control that represent more naturally how humans reason about privacy. Furthermore, these access controls can deal with specific issues that social network services present. Specifically, users often share information that concerns many people, especially other members of the social network. In such situations, two or more people can have conflicting privacy preferences; thus, an appropriate sharing policy may not be apparent. These situations are usually identified as multiuser privacy scenarios.

Since relationship based access controls are complex for the average social network user, service providers have not adopted them. Therefore, to enable the implementation of such access controls in current social networks, tools and mechanisms that facilitate their use must be provided. To that aim, this thesis makes five contributions: (1) a review of related research on privacy management on social networks that identifies pressing challenges in the field, (2) BFF, a tool for eliciting automatically tie strength and user communities, (3) a new access control that employs communities, individual identifiers, tie strength, and content tags, (4) a novel model for representing and reasoning about multiuser privacy scenarios, employing three types of features: contextual factors, user preferences, and user arguments; and, (5) Muppet, a tool that recommends sharing policies in multiuser privacy scenarios.

The contributions of this thesis employ techniques such as artificial intelligence, machine learning, data mining, and crowdsourcing. The contributions are validated by means of studies with human participants. Specifically, three studies with 38, 50, and 988 participants provide the foundational data that is employed to evaluate each contribution. The results show that access control models for social network services can be enhanced by means of tools that automatize privacy related tasks (BFF and Muppet), and formal models that capture accurately how humans reason about information disclosure.

Contents

Part I

Introduction and Objectives

CHAPTER

1

Introduction

Social Network Services (SNSs), such as Facebook or Twitter, appeared more than a decade ago and redefined the way people interact with others on the Internet. With millions of users and millions of photos, comments, and videos uploaded to their servers every day, SNSs have become one of the most prominent services used by people all over the world. Users employ SNSs because the perceive benefits in social capital and opportunities [38]. They can form new friendships, maintain contact with friends and acquaintances, and, in general, be connected with people they care about. However, these benefits also come with a cost. Managing privacy on SNSs is usually a burden that many users do not want to handle, or they are just not capable of [76]. This may lead to privacy breaches that can affect the daily lives of SNS users. For example, nowadays, it is common to find news on the media about people who lost their job because of a post on Facebook.

Since the beginning, SNSs have been of interest in the privacy research field. Early works, such as Lipford et al. [94], found that users were not aware of the dangers of disclosing sensible information on SNS and left, in many cases, their profiles open to everyone. Although more recent research found out that users are no longer that reckless in their privacy management [140], people still struggle to properly control how their information flows on SNSs [42].

SNS providers are aware of the privacy concerns of their current users and they realize that these concerns can act as a deterrent for potential new users. Hence, privacy controls on SNSs have been maturing during the last few years. In general, SNSs employ very simple privacy models that aim at requiring the least amount of effort from users. These models sacrifice complexity in favor of simple configuration. Currently, privacy management on SNSs is based on contact grouping (e.g., Facebook friend lists and Google+ friend circles). To facilitate the definition of groups, SNSs offer graphical tools and predefined groups, such as friends of friends. Additionally, SNSs have lately increased the visibility of privacy configuration, and many SNSs remind periodically users about the importance of proper privacy settings.

A number of research works propose Relationship Based Access Controls (ReBAC) for privacy management on social networks [18, 22, 46]. ReBAC is a paradigm that provides mechanisms for disclosure control based on interpersonal relationships, and they represent more naturally how humans decide what to share and with whom. In contrast to traditional access controls where permissions are granted based on users and their roles, ReBAC models employ social elements such as relationship between the owner and potential viewers (e.g., this photo can only be seen by friends). ReBAC models present rich relationship features such as tie strength and asymmetrical relationships, and propose advanced functionality like sharing policy interoperability. However, ReBAC models proposed in the related literature (or similar ones) have not been implemented on commercial SNSs. One of the reasons why SNSs have not adopted such models is that, even though these models present a more powerful framework for privacy management, SNS providers must weight the trade-off between privacy preservation and usability. This thesis aims at closing the gap between formal ReBAC models and those that are currently in use on SNSs. To achieve this goal, this thesis proposes mechanisms that automatize burdensome privacy related tasks and novel models that capture how users reason about information disclosure. To develop and evaluate the contributions presented in this thesis, a variety of techniques are employed which include machine learning, artificial intelligence, and crowdsourcing.

1.1 Objectives

As explained above, the main goal of this thesis work is to facilitate the implementation of formal ReBAC models in commercially successful SNSs. This goal entails a number of challenges: (1) defining the requirements that a ReBAC model must fulfill, (2) new additions and modifications of current privacy models have to be made considering their effects on the usability of those models, and (3) all proposed methods, models, and tools have to be evaluated with data collected from real users. Considering these challenges, the main goal of this thesis is pursued by means of automatization of privacy related tasks and the definition of ReBAC models that are understandable by users and capture the way they consider information disclosure. The main objective can be subdivided in the following, below detailed, sub-objectives:

1. Reviewing thoroughly related literature on privacy management on social media. This enables the author to find the most pressing research challenges in the field.

2. Based on the challenges found as a result of the first sub-objective, specification of a list of requirements that privacy management systems and ReBAC models for SNSs should provide to enable users to manage accurately their information disclosure.

3. Development of tools that help users automatize totally or partially the

specification of information that new ReBAC models for SNSs will require.

4. Development and evaluation of ReBAC model prototypes that fulfill the requirements identified previously.

5. Definition of an abstract model for representing and reasoning about multiuser privacy scenarios.

6. Development of a privacy recommender that, based on the model previously defined, suggests sharing policies in multiuser privacy scenarios.

1.2 Contributions

To achieve the main objective and its sub-objectives, this thesis works presents the following contributions.

1.2.1 Definition of Open Challenges in Access Controls for SNSs

To define the specific advances that are needed to offer rich ReBAC models to SNS users, first, this thesis dissertation presents a review of current research on the topic of privacy management on SNSs. This review classifies current research into five categories: (1) content type management, (2) co-privacy, (3) modeling human relationships on SNSs, (4) recommender systems, and, (5) improving privacy settings understanding

The review examines and summarizes papers in each category. The summary of each paper highlights its goals and proposals, and reviews the method employed by the authors to evaluate those proposals. Moreover, reviewed research papers are linked by means of conceptual maps that help readers understand, in a visual way, how research on privacy on SNSs is organized.

After summarizing papers in each category, the review lists a number of open challenges that researchers should address in order to improve privacy management on SNSs. Specifically, the rest of this thesis work focuses on three of these challenges: (1) tie strength automatic inference, (2) ReBAC and content type, and, (3) multiuser privacy management.

1.2.2 BFF

Research in social media has found that one of the the most important factors that users consider when disclosing information is the strength and type of the relationships they have with other individuals and the communities in which they are involved [156]. This idea was introduced in the 70's by Granovetter [56]. In his work, Granovetter describes two different types of ties: *strong* and *weak*. On the one hand, strong ties usually include relationships such as family and close friends. On the other hand, weak ties may refer to coworkers or less trusted friends. Moreover, in the contextual integrity framework described by Nissenbaum [109], it is pointed out that distinctive relationships, for example individual to spouse, boss, friend, colleague, and so on, are partially defined by distinctive patterns of information sharing. Based on this scientific evidence, it is reasonable to assume that a richer relationship definition that includes tie strength would help users manage their privacy on SNSs. However, given the large number of contacts that users can have and the number of communities they may be involved on [117], it is not realistic to assume that users are able to specify this information without the whole eliciting process becoming confusing and time consuming. Therefore, to make the use of tie strength in access controls viable, users require tools that automatize this process. To that aim, this thesis presents the tool BFF. This tool is able of inferring tie strength values and communities employing information that is commonly available on SNSs. This information includes elements such as the number of common contacts, number of exchanged messages, and the duration of the relationship on the SNS. BFF works as a recommender, thus, the information inferred by BFF is presented to the users as a suggestion which they can accept, modify, or dismiss entirely. This thesis reports on a study with 38 users that tested an implementation of BFF on Facebook. The results show that BFF is able to infer accurately tie strength values and the composition of communities.

1.2.3 Evaluation of Tie Strength and Tags as Attributes for Access Controls

Once tie strength elicitation is not a burdensome task for users, the possibility of including tie strength as an attribute in a ReBAC model becomes plausible. Furthermore, related literature suggests the use of content tags (i.e., labels that identify the content of the item being shared, for example, *selfie* photo) in access controls to facilitate privacy management [83, 163]. Nonetheless, it is necessary to evaluate the benefits and possible disadvantages of the inclusion of both attributes. To justify their inclusion, users employing tie strength or tags when defining sharing policies should be able to define those policies in a simpler way and they should fully understand the implications of their privacy configuration. To perform this evaluation, first, this thesis presents a preliminary study employing real data from users. This study shows that tie strength and tags are promising attributes for access controls on SNSs. Then, three prototypes of access controls that employ different combinations of attributes (tag, tie strength, group, and individual identifier) are developed. These prototypes are tested in a new study. The prototypes are evaluated through a number of quantitative metrics which measure the usefulness and performance of each attribute and their combinations. Further, a qualitative analysis shows that users prefer access controls that employ tie strength and tags. In general, the results indicate that tie strength and tags could be useful additions to current ReBAC models for SNSs. Although, users would need assistance while using these two attributes.

1.2.4 Modeling Multiuser Privacy Scenarios

One of the particularities of the information shared on SNSs is that, often, it involves more than one user. A natural example is a picture or video showing a group of people. Many SNSs enable users to connect the information they upload to other users so that the connected users can be notified of the uploaded information. Since the information shared varies depending on the SNS, these connections can take different forms, e.g., tags on a picture uploaded to Instagram or mentions in a tweet. Suppose Alice uploads a picture from last weekend's party in which she appears together with her friend Bob, and tags Bob in the picture. When these connections are created, the other users are linked to the uploaded information. Usually, a connection implies that the profile of the user can be accessed from the uploaded data or some personal information is shown in conjunction with the uploaded data. Although connections between information and users are widely employed by SNSs, they can pose a privacy threat. For example, Bob may find that the picture Alice uploaded is sensitive. However, Bob has no control over uploading that picture, and Alice's action can threaten Bob's privacy by revealing information about him from one setting or context into another. In this work, a situation such as this is identified as a *multiuser privacy scenario* or, for brevity, *multiuser scenario*.

Currently, SNSs do not provide mechanisms to handle multiuser scenarios [42]. Thus, a user who did not upload a piece of information concerning him must deal with the privacy settings chosen by the uploader; at best, the user can remove the connection that links him to the shared information, but the information itself remains nevertheless. An ideal solution in a multiuser scenario is to respect each user's privacy. However, often such a solution may not be viable since the preferences of the users involved may conflict.

Evidence from self-reported data [6, 87, 159] suggests that, when dealing with a multiuser scenario, users entertain the explanations provided by others and that the optimal solution may depend upon the particular context and reasons behind users' preferences. Following this idea, this thesis presents a formal model that considers three types of factors that potentially influence a privacy decision: the scenario's *context*, users' *preferences*, and their *arguments* about those preferences.

To evaluate the proposed model, this thesis reports on a study with 988 participants. This study is based on surveys that present hypothetical scenarios to participants where a group of users have to determine the appropriate sharing policy to solve a multiuser conflict. The results indicate that all features considered by the model have an influence on the final sharing policy.

1.2.5 Muppet

Finally, the last contribution of the thesis is Muppet, a recommender for multiuser scenarios. Few recent works focus on helping users deal with multiuser scenarios. Squicciarini et al. [133] propose an auction-based framework to help users reach an agreement. Other approaches elicit sharing policies based on fixed [147] or variable [71] preference aggregation methods. Hu et al. [72] describe a game-theoretic mechanism for multiparty access control. These works base their recommendations on the sharing preferences of the users and require users specifying all the information. Muppet, instead, takes into account elements such as the context, characteristics of users (individually and as a group), and the relationship among them, and can generate a recommendation even when some information is unknown.

To avoid the common problems of recommenders known as cold start, this thesis presents a bootstrapping approach that enables Muppet to be functional off-the-shelf.

1.3 Structure of the Thesis

Considering the motivations and objectives of this thesis, the rest of the document is organized as follows:

1. **Part I. Introduction and Objectives:** This part presents the motivation and the

goals of the thesis as well as the structure of this document.

2. **Part II. Selected Papers:** This presents a selection of the most relevant articles supporting this theses which were published in conferences and journals.

3. **Part III. Discussion:** This final part discusses the results obtained in the published works and possible paths for future work, and presents some concluding remarks.

1.4 Publication List

In this section, all the international publications related to this thesis are listed. They have been classified according to their type (journals or international conferences) as well as whether they are listed in JCR or in CORE, respectively. Those publications which have been included in this document are marked with (*).

- Journals listed in JCR:
  - (*) R. L. Fogués, P. K. Murukannaiah, J. M. Such and M. P. Singh. *Sharing Policies in Multiuser Privacy Scenarios: Incorporating Context, Preferences, and Arguments in Decision Making.* **ACM Transactions on Computer-Human Interaction** In press, 2017 **Impact Factor: 1.293**
  - (*) R. L. Fogués, J.M. Such, A. Espinosa and A. García-Fornes. *Open Challenges in Relationship-Based Privacy Mechanisms for Social Network Services.* **International Journal of Human–Computer Interaction** Vol. 31 N. 5 pp. 350-370. (2015) **Impact Factor: 1.260**
    DOI: 10.1080/10447318.2014.1001300
    http://dx.doi.org/10.1080/10447318.2014.1001300
  - (*) R. L. Fogués, J.M. Such, A. Espinosa and A. García-Fornes. *BFF: A tool for eliciting tie strength and user communities in social networking services.* **Information Systems Frontiers** Vol. 16 N. 2 pp. 225-237 . (2014) **Impact**

  Factor: 1.077 · DOI: 10.1007/s10796-013-9453-6
  http://dx.doi.org/10.1007/s10796-013-9453-6

  - (*) R. L. Fogués, J.M. Such, A. Espinosa and A. García-Fornes. *Tie and Tag: A Study of Tie Strength and Tags for Photo Sharing*. **Plos One** Submission under review **Impact Factor: 3.057**
  - (*) R. L. Fogués, P. K. Murukannaiah, J. M. Such and M. P. Singh. *SoSharP: Recommending Sharing Policies in Multiuser Privacy Scenarios*. **IEEE Internet Computing** To be published **Impact Factor: 2.296**
  - R. L. Fogués, J.M. Such, J.M. Alberola, A. Espinosa and A. García-Fornes. *Supporting Dynamicity in Emergency Response Applications*. **Computing and Informatics** Vol. 33 N. 6 pp. 1288-1311. (2014) **Impact Factor: 0.504** · ISSN: 1335-9150
    http://www.cai.sk/ojs/index.php/cai/article/view/2815/676

- International conferences listed in CORE:
  - (*) R. L. Fogués, J.M. Such, A. Espinosa and A. García-Fornes. *Exploring the Viability of Tie Strength and Tags in Access Controls for Photo Sharing*. **32nd Symposium on Applied Computing**, ACM 2017, Marrakesh, Morocco, April. In press. 2017. DOI: 10.1145/3019612.3019909. **CORE ERA2014 Rank: B**.
    http://dx.doi.org/10.1145/3019612.3019909

- Other international conferences:
  - R. L. Fogués, P. K. Murukannaiah, J.M. Such, A. Espinosa, A. García-Fornes and M. P. Singh. *Argumentation for multi-party privacy management*. **The Second International Workshop on Agents and CyberSecurity (ACySe)**, Istanbul, Turkey. 2015.
    http://nms.kcl.ac.uk/jose.such/acyse2015-proceedings/ACySe2015_submission_Fogues.pdf
  - R. L. Fogués, J.M. Such, A. Espinosa and A. García-Fornes. *A Tool for Retrieving Meaningful Privacy Information from Social Networks*. **ITMAS**

2012: Infrastructures and Tools for Multiagent Systems, Valencia, Spain. 2012. ISBN: 978-84-8363-850-7
https://riunet.upv.es/bitstream/handle/10251/16889/ITMAS%202012%20.pdf

- R. L. Fogués, J.M. Alberola, J.M. Such, A. Espinosa and A. García-Fornes. *Towards dynamic agent interaction support in open multiagent systems.* **Artificial Intelligence Research and Development: Proceedings of the 13th International Conference of the Catalan Association for Artificial Intelligence** Vol. 220 N. 5 pp. 89-97. (2010) DOI: 10.3233/978-1-60750-643-0-89
http://dl.acm.org/citation.cfm?id=1893268.1893282

The publications selected to be included in the Part II of this document are the most relevant and closely related to this research work, regarding the objectives set in Section 1.1. The contributions of the author of this thesis in each paper are specified employing the CRediT taxonomy of author contributions [15].

Chapter 2 (previously published in [42]) presents a review of related work on the field of privacy on SNS. This paper starts introducing a list of privacy threats that can affect SNS user and what requirements privacy mechanisms should fulfill to prevent this threats. Then, it reviews current approaches and analyze to what extent they cover the requirements. Finally, it presents a number of challenges that future research work should address in order to enable SNS users manage their privacy easily and efficiently. The contributions of the author in this paper are: conceptualization, investigation, visualization, and writing.

Chapter 3 (previously published in [41]) introduces BFF that automatically classifies the friends of a user in communities and assigns a value to the strength of the relationship ties to each one. To provide predictions, BFF exploits different data available on SNSs. Such data includes elements as shared items between two users, and number of common friends. To test BFF, the chapter presents an evaluation with thirty-eight human subjects. The results point out that BFF is able of providing accurate predictions employing a rather low number of variables. The contributions in this paper are: conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, and writing.

Chapter 4 (previously published in [43]) studies the viability of tie strength and tags as attributes for access controls on SNSs. To this aim, the chapter defines three metrics that enable researchers to evaluate the ease of use and complexity of access controls. Then, employing the data collected in a study with human subjects, fifteen access controls, each one with a different combination of attributes, are evaluated. The contributions in this paper are: conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, and writing.

Chapter 5 (submission currently under review at Plos One) describes an experiment with forty-eight participants using access controls that include tie strength and tags (separately and simultaneously) together with groups and individuals. The results of the study are analyzed employing a number quantitative metrics. Additionally, a qualitative analysis is also performed on the collected data. The results show that, users prefer access controls that employ tags and tie strength. Moreover, users employ these two attributes extensively when they are available in the access control. However, the results also indicate that users make more mistakes in terms of sharing policy correctness when tie strength or tags are employed. This points out that users need assistance when employing these two new attributes. The contributions are: conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, and writing.

Chapter 6 (previously published in [44]) presents a formal model for representing and reasoning about multiuser scenarios, employing three types of features: contextual factors, user preferences, and user arguments. The Chapter also reports on a study that involved a survey of 988 Amazon MTurk users about a variety of multiuser scenarios and the optimal sharing policy for each scenario. The evaluation of the participants' responses reveals that contextual factors, user preferences, and arguments influence the optimal sharing policy in a multiuser scenario. Finally, the responses obtained in the study are employed to develop and evaluate an inference model that predicts the optimal sharing policy for a multiuser scenario. The contributions are: conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, and writing.

Chapter 7 (published in [45]) introduces Muppet, a sharing policy recommender tool for muiltiuser scenarios. Muppet recommends a sharing policy considering the (1) contextual features defining the scenario, (2) characteristics of the users involved, (3) their preferences, and (4) group characteristics. Muppet works incrementally and asks for users' input only when required. The Chapter also addresses the typical problem of recommenders known as cold start, i.e., the issue that a system cannot draw inferences when it has not gathered sufficient information. To this aim, a bootstrapping technique employing crowdsourced training data is presented. The contributions are: conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, and writing.

1.5 Research Projects

The research work presented in this PhD thesis was carried out in the context of the following research projects:

- **Privacidad en entornos sociales educativos durante la infancia y la adolescencia**
  - Funder: Ministerio de Ciencia e Innovación TIN2014-55206R
  - Lead Applicant: Ana María García-Fornes
  - Years: 2015 - 2017
- **Interacción multiagente para planificación**
  - Funder: Ministerio de economía, industria y competitividad TIN2008-04446/TIN

  - Lead Applicant: Eva Onaindia De La Rivaherrera
  - Years: 2012 - 2013

- **Advances on agreement technologies for computational entities**
  - Funder: Generalitat Valenciana Prometeo/2008/051
  - Lead Applicant: Vicente Juan Botti Navarro
  - Years: 2010 - 2012

- **MAGENTIX II: Una Plataforma para Sistemas Multiagente Abiertos**
  - Funder: Ministerio de Ciencia e Innovación TIN2008-04446/TIN
  - Lead Applicant: Ana María García-Fornes
  - Years: 2009 - 2011

Part II

Selected Papers

CHAPTER 2

Open Challenges in Relationship-Based Privacy Mechanisms for Social Network Services

AUTHORS:

RICARD L. FOGUES*, JOSE M. SUCH†, AGUSTIN ESPINOSA* AND ANA GARCIA-FORNES*

{rilopez,aespinosa,agarcia}@dsic.upv.es, jose.such@kcl.ac.uk

*DEPARTAMENTO DE SISTEMAS INFORMÁTICOS Y COMPUTACIÓN
UNIVERSIDAD POLITÉCNICA DE VALENCIA, SPAIN
†KING'S COLLEGE DEPARTMENT OF INFORMATICS LONDON, UK

Abstract

Social networking services (SNSs) such as Facebook or Twitter have experienced an explosive growth during the few past years. Millions of users have created their profiles on these services because they experience great benefits in terms of friendship. SNSs can help people to maintain their friendships, organize their social lives, start new friendships, or meet others that share their hobbies and interests. However, all these benefits can be eclipsed by the privacy hazards that affect people in SNSs. People expose intimate information of their lives on SNSs, and this information affects the way others think about them. It is crucial that users be able to control how their information is distributed through the SNSs and decide who can access it. This paper presents a list of privacy threats that can affect SNS users, and what requirements privacy mechanisms should fulfill to prevent this threats. Then, we review current approaches and analyze to what extent they cover the requirements.

2.1 Introduction

The advent of the Web 2.0 has supposed a revolution in how users interact with Web technologies. Social network services (SNS) are some of the most successful applications of this revolution [26]. Facebook with more than 900 million active users[1], Twitter with more than 500 million registered members[2], and Qzone with more than 51 millions of users are some of the biggest SNSs. The impact of these services on society, especially on young people, is unquestionable.

Privacy problems associated with digital communication and network technologies have been a major concern among Internet users over the past decade [162]. The emergence of social networks has even increased these concerns. People register to these SNSs and share images, videos, and thoughts because they perceive a great

[1] Facebook statistics `http://newsroom.fb.com/`

[2] Twitter To Surpass 500 Million Registered Users On Wednesday. `http://www.mediabistro.com/alltwitter/500-million-registered-users`

payoff in terms of friendship, jobs, and other opportunities [38]. The popularity of SNSs attracts not only faithful users but third parties with adverse interest [3]. If we consider the huge amount of private information uploaded to those SNSs and the persistence of it in the social networks, the privacy of SNS users can be threatened [58]. Recent cases show that on-line thieves, stalkers, and bullies take advantage of the information available on SNSs and use it for purposes that were not the initially intended ones [65].

There are several definitions of privacy in the related literature. In the context of this survey, we use the definition of Alan Westin, who defined privacy as ”the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated“ [155]. This definition implies that SNSs have to offer their users mechanisms that allow them to decide how their information is disclosed. Current SNSs have taken steps towards this objective, but there still exist several problems that make users feel they have lost control of their information and how it is shared among the SNS [68, 139]. Users demand better privacy mechanisms, with richer and finer-grained privacy policies that take into account the way SNS users share information and interact with others. Moreover, privacy controls for these new access controls have to be easy to use, offering automatic suggestions and learning from the behavior of the users.

This article reviews studies that advance the state of the art on interpersonal privacy management in social networks, as well as studies that improve the usability of such mechanisms. Understanding how humans share and manage their friendships on SNSs is crucial so that researchers can adapt their models and methods to cope with the users' needs and expectations. Studies are classified according to the type of privacy requirement they address. Each study is impartially presented and reviewed, showing the strengths and weaknesses of the proposals made in the study. SNS became a global phenomenon around 2006-07 and immediately attracted the attention of researchers [12]. We consider that the first meaningful research works on interpersonal privacy appeared shortly after, thus, we only survey papers published from 2008 onwards. The final objective of this paper is to survey research in the field of interpersonal privacy in social networks as well as to promote and encourage future research and advances that overcome the current challenges that exist in this field.

2.1.1 Privacy and Social Networks

As pointed out more than a century ago by Warren and Brandeis [152], disclosure of private information and the misuse of it can damage people's feelings and cause considerable damage in people's lives. In SNS where intimate information of the users is managed, privacy is of paramount importance. A research of Gross and Acquisti [58] in the early days of Facebook showed that the majority of users were unconcerned about privacy risks. They tended to use default privacy configurations and personal data were generously provided. More recent studies, like the one from Boyd and Hargittai [13], show that the privacy awareness of SNS users has increased lately. The widespread media attention on SNS and on situations where the leakage of personal information of SNS users affected their lives has positively influenced the way SNS users manage their privacy [116]. Moreover, as people get used to employ SNS, they can control more effectively how their information is disclosed [127]. Nevertheless, the high number of privacy risks that affect SNS users leaves room for improvement in this field of study.

The most important SNS users' privacy concerns are: identity theft [7],unauthorized access [116], misuse of personal information and stalking [58, 14, 124], and profiling [62]. Many of these threats affect affect levels of SNS privacy mechanisms that are not affected by human factors. Identity theft and unauthorized access are related to access control enforcement. For example, unauthorized access can occur if the authentication mechanisms of the SNS are not good enough or if the communication between the user and the SNS is not properly encrypted. Profiling is a threat when the party which owns the information on the SNS is not trustworthy. A typical case of profiling occurs when the party that manages the SNS sells incorrectly anonymized information to third parties that use it for marketing purposes. A number of surveys that present works that deal with these threats has been recently published [78, 167].

In this study we focus on the threat that is related to interpersonal privacy: the misuse of personal information. This threat refers to the possibility of a malicious dissemination of previously collected information. For instance, users may face blackmailing situations when embarrassing data is collected from a SNS by a third party. In the context of SNSs, misuse of personal information usually occurs when users disclose inappropriate information due to a negligence during the configuration of their privacy settings or ignorance about how privacy is managed on the SNS. Many times inappropriate information is leaked to people inside the user's social network. Johnson [76] classified this threat as the *insider threat* and found out that this is one of the most worrisome threats for SNS users. Such privacy leak can occur because SNS privacy controls are not usable enough or because interpersonal relationships are not well represented on the SNS.

It has been acknowledged that in order to properly minimize misuse of personal information, a new privacy mechanism is needed [51, 166]. In the next section, we detail the requirements for such a new privacy mechanism.

2.1.2 Requirements for a Social Network Service Privacy Mechanism

Any privacy mechanism has at its base an access control. Access controls dictate how permissions are given, what elements can be private, how access rules are defined, and so on. Access control models of current SNSs tend to be very simplistic. Nonetheless, recent improvements in facebook-like SNSs have enhanced the access control models. For example, now it is possible to define policies to deny access to groups of users, instead of individuals. Some SNSs allow the possibility to express a social distance of contacts that have access to the resource, for example, friends of friends (two hops), friends of friends of friends (three hops), and so on. Another addition that only a few SNSs have added is the possibility to choose the amount of information from a friend that we want to receive (Facebook). However, these models still lack key elements. One of the most important is the lack of diversity in the type of relationships. Most SNSs only employ ”friend” as the only type of possible relationship. This lack of classification of contacts leads to privacy leaks to other members inside the social network. Gates [51] identifies the following requirements that an access control model for a SNS must fulfill:

- **Relationship-based**: People base their decision of sharing information on their relationship with others. Moreover, the properties of the relationship also affect the way people disclose their personal information. In social psychology, it is generally accepted that one discloses more of his/her personal information to someone in a strong relationship [36]. Hence, a control access model that tries to reflect the way people disclose and share in real life should be based on relationships.

- **Fine-grained**: The access control has to allow users to define access policies for single items. If the access control is available in a fine-grained format the privacy policies can be more flexible and they can express the user’s preferences exactly. For example, a user should be able to define privacy policies for specific photos, individual blog entries, or even some words or phrases of a comment. In other words, users should be able to decide exactly to what extent others can access their information.

- **Interoperability**: Many SNSs have a specific objective; while Facebook aims to facilitate users contacting their friends, LinkedIn helps users to maintain their professional networks. Facebook and Linkedin have clearly different purposes. Because of this variety of purposes, users may have several multiple accounts in different SNSs, each one for a different social objective. In this scenario, it is highly desirable for access controls to be interoperable and follow the users, so it is not necessary to define an entire new access control for each SNS.

- **Sticky policies**: Besides being interoperable, privacy policies should also follow the data to which they apply. For example, many SNS allow third party applications to access users' data. The privacy preferences assigned to that data should be respected by these third parties and in whatever context it might travel to. This idea was introduced by Karjoth et al. [82].

In the related literature, we have also identified additional requirements that play a crucial role in developing successful access control models for SNSs:

- **Content Type Management**: SNS enable users to share a variety of different pieces of information: photos, videos, comments, events, hobbies, and so on. Besides the miscellany in the format of the information, its content also matters when deciding who has access and who has not [61]. Flickr[3] employs tags so users can classify their pictures according to their type. A similar approach could be used so users could define permissions based on the type of the content.

- **Co-privacy**: SNS users like to upload items to their profiles, such as photographs and videos, where other users are depicted. Specially, SNSs that focus on helping users to maintain their friend relationships encourage users to upload publications of this kind. Items of this type can raise several privacy concerns. While the owner of the item is in charge of assigning a privacy policy to it, the other users related to the item can be affected if the privacy policy is not appropriate for their interests. It is possible to infer a great amount of information about an individual from information leaks that occur due to shared items and privacy preference conflicts [147]. Current access models do not consider these situations; thus, users are forced to use strategies like untagging, asking the owner to remove the photo or, in the most extreme situations, removing friendship links. Access controls should consider co-privacy management and offer mechanisms that allow every user involved in a single item to express their privacy preference so that the resulting privacy policy applied to that item maximizes the utility for everyone.

[3] www.flickr.com

An access control acts as the base for a privacy mechanism, however, they require other elements to be functional. It is not realistic to assume that SNS users can understand access control models and use them intuitively. Powerful privacy models are useless if they lack usability and are not understood by the people that will use them [27]. Moreover, as pointed out by Brandtzæg et al. [16], in order for SNSs to be successful their users have to be able to easily control the disclosure of their information. If users have privacy concerns, they share less content, thus, losing social opportunities that SNSs offer. Users need tools that guide them through the process of setting their privacy preferences. Users also require mechanisms that help them to understand their current privacy preferences and how their information is disseminated among other SNS users. According to related literature, a privacy mechanism for SNSs should fulfill these requirements:

- **Automatic relationship inferring**: If the access control has to be based on social relationships, these have to be accurately defined. SNS users tend to have a high number of friends. For example, according to the Facebook statistics, the average number of friends in that social network is 130. Hence, classifying every contact in a social network can represent a burden on the user. Privacy mechanisms should have the capacity to automatically infer the type of a relationship and make the whole process of friend classification easy and fast.
- **Privacy setting recommendation**: While privacy is paramount on SNSs, users are focused on enjoying the functionality that these offer. For many users privacy settings represent a burden, for others privacy settings are difficult to manage and understand. Recommender tools can help users to set properly their privacy settings. While recommenders can help reduce the user's burden, they are rarely perfectly accurate. Thus, it is important for the user to be able to view, understand, and modify the recommended policy before it is applied
- **Privacy understandability**: Access controls can be complex and daunting for SNS users. Average SNS users do not have expertise on security, thus, it is difficult for them to accurately evaluate how their information is disclosed through the SNS

[103]. Users require proper interfaces that show them how their privacy policies dictate their self-disclosure.

- **Self-presentation management**: In the beginning, social media was focused on establishing or maintaining friendship relationships through a digital channel. However, social media have increased the number of services offered and now users expect more benefits than friendship alone [38]. Some examples of social media use are to obtain fame[4], or for commercial brands to publicize their products, acquire recognition, and maintain contact with their customers[5]. In a nutshell, and as pointed out by Kairam et al. [81], social media users utilize the product to successfully tailor self-presentations for various parts of their network through selective information sharing. Self presentation is achieved by carefully tailoring self-disclosure. Privacy controls should help users to maintain their chosen self presentation on SNSs.

In the following sections of this paper, we will review studies that aim to totally or partially cover the requirements previously listed. First, Section 2.2 starts reviewing formal relationship-based access control models. Section 2.3 presents some prototypes of access controls that take into account the content type of the information being shared. Section 2.4 is centered on papers that deal with co-privacy. Section 2.5 presents papers that aim to accurately model and infer human relationships on SNSs. Section 2.6 presents works that propose privacy policy recommenders. Section 2.7 reviews papers that present methods to enhance the understandability of privacy settings. The remaining requirements (interoperability, sticky policies, and self-presentation management) are treated as open challenges and are covered further in section 2.8. Figure 2.1 shows a conceptual map that presents the Sections that cover each requirement.

[4] `http://www.wltx.com/news/tech/article/233056/378/Survey-Social-Media-Draws-Young-Fame-Seekers`

[5] `http://www.usatoday.com/story/money/personalfinance/2013/04/16/small-business--social-media-facebook/2075123/`

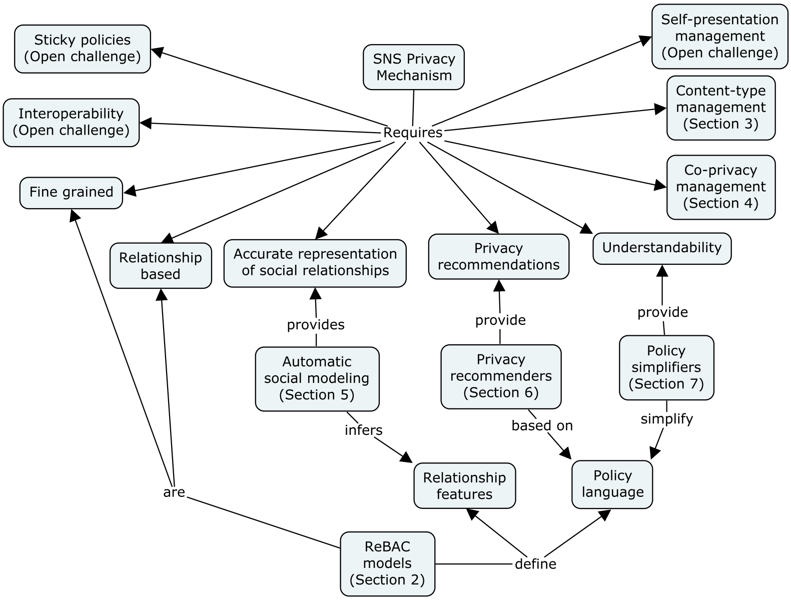

Sticky policies (Open challenge)
SNS Privacy Mechanism
Self-presentation management (Open challenge)
Interoperability (Open challenge)
Requires
Content-type management (Section 3)
Fine grained
Co-privacy management (Section 4)
Relationship based
Accurate representation of social relationships
Privacy recommendations
Understandability
provides
provide
provide
Automatic social modeling (Section 5)
Privacy recommenders (Section 6)
Policy simplifiers (Section 7)
infers
based on
simplify
are
Relationship features
Policy language
ReBAC models (Section 2)
define

Figure 2.1: Conceptual map of requirements for a privacy mechanism for SNSs

2.2 Relationship-based Access Control Models

This section reviews Relationship-based Access Control (ReBAC) models proposed for SNSs. The ReBAC paradigm provides users with better mechanisms for disclosure control, and they represent more naturally how humans decide what to share and with whom. All models studied in this section fulfill the two first requirements of access controls for SNSs. Specifically, all of them are based on relationships and their privacy policies are fine grained. However, the models differ in other features, table 2.1 shows a comparison of them. In the table, the features of the reviewed models have been divided into two subgroups.

The first group contains the properties that the models consider for the relationships. *Multiple relationship type* refers to the possibility that the model manages different kinds of relationship; for example, a model could allow the differentiation between a friend and a family relationship. Another concept used to differentiate relationships, which is further explained in section 2.5, is tie strength. Some models allow the possibility of specifying a numerical value for the strength of the relationship. *Directional relationship* alludes to the possibility of defining asymmetric relationship; for example, user *A* can be related to user *B* but that does not imply that the opposite relationship exists. Finally, the last two features, *user-to-user relationship* and *user-to-resource relationship*, describe whether the models consider relationships among users and among users and items. For example, a user can somehow be related to a movie, which is not a user but a resource.

The second group of features includes those that affect the policy specification language of each model. As explained in the previous section, ReBAC models have to be fine-grained; the *policy individualization* refers to this characteristic. The following two features, *regular expression language* and *based on ontology*, specify whether the policy language uses any of these techniques. When specifying a privacy policy, a common resource is to define a maximum social distance, which is the number of hops between the owner of the resource and the accessor. *Arbitrary social distance* identifies what models allow this possibility. Besides the social distance, *social path specification* defines the type of the hops of that distance; for example, the family members of my friends represent a social distance of two, where the first hop is friends and the second family members.

Fong et al. [47] proposed a formal algebraic model for facebook-style social networks. The authors created this algebraic model to reflect how facebook-style SNSs model their access control. Even though this model cannot be classified as a true ReBAC, we considered this paper in this review since it formally shows the limitations of current SNS access control models. The model divides the authorization of access to a resource into two stages. Stage 1 is to reach the profile of the resource's owner and Stage 2 is to access the resource. The model allows the definition of authorization policies for each stage independently. Since policy language considers the network topology properties, the model allows complex policies that are beyond what facebook-like SNSs offer (friends, friends of friends, no

	[47]	[46], [18]	[22], [21]	[23]	[25]
Relationship features					
Multiple relationship types		⋆	⋆	⋆	⋆
Tie-strength			⋆	⋆	
Directional relationship		⋆			⋆
User-to-user relationship	⋆	⋆	⋆	⋆	⋆
User-to-resource relationship				⋆	
Policy language features					
Policy individualization	⋆	⋆	⋆	⋆	⋆
Regular expression language					⋆
Based on ontology				⋆	
Arbitrary social distance	⋆	⋆	⋆	⋆	⋆
Social path specification		⋆	⋆	⋆	⋆

Table 2.1: Comparison of ReBAC models.

one, and public). Topology-based policies include: degree of separation, *k* common friends, *k* clique, trusted referral, and stranger. Since this work aims to model the access control of facebook-like SNSs, it also has the same limitations. Moreover, since users' profiles and resources are not treated the same, the authorization process is divided into two steps. This division can lead to unnecessary complexity of access policies.

Fong [46] proposes an access model for social networks based on relationships. In contrast to [47], this work uses social contexts and allows a generalization of relationships types (e.g. parent-child, employer-employee, etc). Social context is another dimension of relationships; some relationships have a different meaning depending on the context or they only exist in a given context. For example, a physician who is my treating physician in one medical case may very well be a consulting expert in a different medical case of mine. As a result, the physician may enjoy a different level of access in each case. This access model considers that, for each relationship, the social network defines its inverse. For example, if a social network has the relationships *parent* and *employer*, it must also contain the relationships *child* and *employee*, which are the inverse of *parent* and *employer* respectively. When a resource is being accessed, the evaluation of the authorization of access for that accessor is done based on an active context. This concept captures how people are willing to disclose different information depending on the context. The policy language defined by Fong allows the specification of unlimited sequences of relationships. For example, it is possible to express a policy that allows access to the father of a friend of a friend. This feature is an improvement with respect to [47] where the chain of relationship was restricted to friend and friend of a friend.

Bruns et al. [18] improved the previous work of Fong [46] adding *Hybrid Logic* to the model. The policies are divided into two sub-policies; one sub-policy is defined from the point of view of the owner of the resource and the other one from the point of view of the accessor. The improved model allows more flexible policies; for example, it is possible to grant access to the last four friends, or grant access if at least *n* friends of the owner fulfill a certain requirement.

These three works [47, 46, 18] share the same limitation. They do not consider the strength or intensity of the relationships (i.e., they only consider relationships as a boolean: either a relationship exist or not).

Carminati et al. [22, 21] propose a model that allows the specification of access rules that consider the type of the relationship, its depth, and its intensity. The proposal of Carminati et al. considers a distributed SNS; therefore, principals are in charge of specifying their access rules. The work also proposes a semidecentralized access control enforcement. When a principal wants to access a resource, she has to prove that she fulfills the requirements specified by the owner of the resource. A central and trusted server is responsible for storing all the relationships of the social network users. Thus, whenever the requester needs to prove to the resource owner the existence and the attributes of a given relationship, she requests this trusted server for this information. The policy language proposed by Carminati et al. allows the definition of policies that specify a type of relationship, a maximum depth, and a minimum strength. Policies can have several requirements, all of them have to be satisfied in order to obtain access. Several policies can be defined for a single resource. In this situation, only one of these policies have to be fulfilled in order to obtain access to the resource. One limitation of the policy language is that policies cannot refer to a chain of relationships with different types of relationships. For example, it is not possible to specify a policy that grants access to the parents of the owner's friends.

Carminati et al. [23] propose an access control model based on semantic web technologies. This paper is an extension of the previous paper [22]; the main differences between the two proposals is that this model considers the user-to-resource relationship and it uses semantic technologies for the policy language. The model proposed by Carminati et al. considers the following five important elements of an online social network: (i) profiles, (ii) types of relationships among users (e.g. Bob and Alice are colleagues), (iii) resources, (iv) relationships between users and resources (e.g. Bob appears in a photo owned by Alice), and (v) actions. The use of semantic web technologies allows the model to infer about the relationships among users and resources. For example, it is possible to infer that a close friend is also a friend and anything that is accessible by friend could also be accessible by a close friend. The authors focus their article on the addition of semantic technologies to their previous access control model [22]. However, the authors did not evaluate how the addition of semantic technologies improved their previous work.

Cheng et al. [25] developed a ReBAC model using regular expression notation. Their model defines resources and users as the target of an action. The model permits a high generality of relationship paths in its policy specification, since the notation of the model is regular expression. The paths can be defined as patterns; for example, it is possible to define a policy that grants access to users that are connected to the resource owner by a path that contains at least one friend and a maximum of two coworkers. Since the users can be considered as targets of an action, it is possible to specify polices that hide the profile of the users and do not show them in searches performed by others that do not satisfy the specifications of the policy. Even when the model considers different types of relationships, it does not contemplate a value for the trust or strength of the relationships. This limitation restricts the power of expressiveness of the model because it is not feasible to define a type of relationship for each possible level of tie strength.

The models reviewed in this section are based on human relationships and depend on them to express privacy policies and define how the different elements can be accessed and by whom. These models assume that the social network provides a rich social model that is capable of representing different types of human relationships. Unfortunately, this is not true in current SNSs, as they usually only consider friend as the only type of possible relationship. Section 2.5 reviews studies that model human relationships and propose theoretical models and actual software tools to accurately predict and represent the type of a relationship and its intensity.

2.3 Content Type Management

None of the access control models reviewed above consider the type of the shared information. In other words, users cannot specify privacy policies for different types of content. For example, a user could not define a privacy policy that affects their family photos. A number of studies [61, 83, 163] show that content matters during privacy policies definition. Moreover, mechanisms to classify content, such as tags, improve the usability of privacy mechanisms.

Yeung et al. [163] prototyped the management of privacy for photos that considers content type. To classify the photos, the authors proposed the use of tags, which is the act of assigning descriptive keywords to resources. This is the same method that Flickr, the popular social network for photo sharing, employs so that users can classify their pictures according to their type. Yeung's system is based on OpenID as authentication protocol and the AIR policy language [79] which is based on RDF. The authors only proposed a prototype of the system and did not provide any evaluation.

Hart et al. [61] proposed a mechanism to manage privacy for blogs based on tags. The authors proposed a privacy language called Plog. This language is based on groups and post type. Users can define groups of users manually or group potential viewers by attributes that they all share (e.g., workplace or same school). The main focus of their study is to compare basic privacy policy mechanisms for blogs with a tag-based approach. To this aim, they developed a WordPress plugin and recruited twenty eight participants to evaluate their proposal. The authors did not use real data from the participants, instead, they created artificial data for imaginary users and asked the participants to manage that data as if it was theirs. Thus, they did not examine users' actual preferences. Their results showed that an approach that uses tags is more usable than one that does not.

Klemperer et al. [83] evaluated the usability of an access control based exclusively on category tags. The authors aimed at evaluating whether tags can be used to organize photos and define their privacy at the same time. Besides, they also studied if tags can decrease the number of privacy conflict. This is, privacy policies that are contradictory. Usually, this happens because users have problems building mental models of their privacy. For their study the authors developed an application that allowed participants to tag their pictures and assign privacy policies for their contacts. The authors asked the participants to use personal photos that were not necessarily uploaded to any SNS. Their results showed that the use of tags reduces the number of required privacy policies and also the number of privacy conflicts.

Paradesi et al. [111] propose a framework for information share based entirely on content. When users employ this framework, they assign keywords to their information. Then, they create privacy policies using these keywords. For example, a user can specify a policy that hides his diabetes when an accessor is looking for his medical information. When users are specifying keywords, the framework takes advantage of a semantic enhancement mechanism that automatically search for related terms. The aim of the framework is to protect private information from agents that are external to the SNS. Therefore, it does not employ any relationship concept to define disclosure preferences. Future research should investigate whether a similar approach could be combined with social concepts, such as tie strength, to control the flow of information inside the SNS.

2.4 Co-privacy

As explained before, one of the requirements for ReBAC models is the management of co-privacy or, in other words, the management of the privacy settings of items shared among users of the social network that affect the intimacy of several individuals. An example of a common issue with items of this kind is a photograph with many users tagged in it. The user who took the photo uploads it to her SNS profile and tags every other user that appears in the photo. At this moment, SNSs leave the responsibility of setting a proper privacy setting for the shared item on the hands of the owner. This decision may suppose a threat to the privacy of the other involved users. The proposals reviewed in this section are focused on finding a proper privacy setting for items that involve several users.

Figure 2.2 depicts a conceptual map for all of the reviewed approaches that propose a co-privacy management mechanism for SNSs. As shown in the conceptual map, there are three different approaches for co-privacy management. *Condition preference sensitivity* refers to the possibility of the users involved to express how willing they are to allow violations of their preferences. The *Suggestions* approach is based on the idea that the owner of the item is the sole person responsible for the privacy management of that item; thus, other users are only allowed to suggest privacy configurations. Finally, the approaches that guarantee the *preferences of everybody* allow every user involved in the item to express their privacy preferences. Then a privacy policy is generated from the combination of every privacy preference.

Squicciarini et al. [133] propose collective privacy management based on the Clarke-Tax algorithm and incentives for users. Incentives are credits that are given

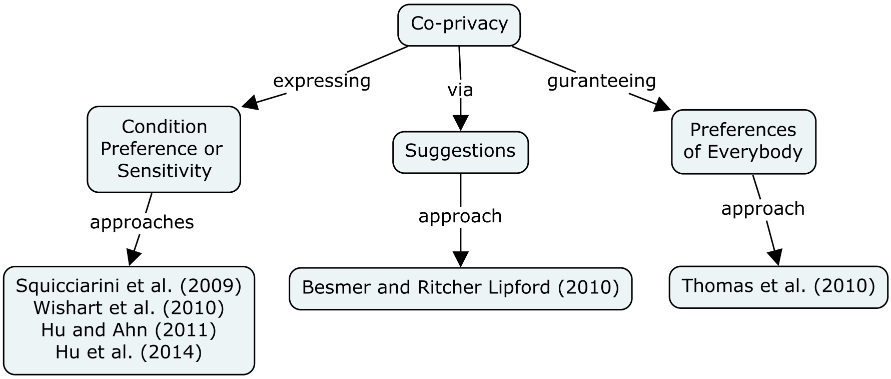


Figure 2.2: Co-privacy Conceptual Map

to users to encourage them to assign co-owners of updated items. When the owner and co-owners specify their privacy preferences for a shared item, each one specifies their privacy preferences assigning a value of credits they are willing to spend in order to apply that policy. The Clarke-Tax algorithm ensures that the privacy policy that maximizes the utility is chosen. The authors created a proof-of-concept application and tested its performance in terms of computing time. The authors propose a novel method that encourages users to participate in the collective managing of privacy by giving them rewards and finds the privacy configuration that has the highest utility according to the co-owners preferences. However, in the study, the privacy policies are very simple and owners only can specify if they prefer the item to remain private (only co-owners have access to it), if it is accessible to co-owners' friends, or if it is designated as public. These simplified privacy preferences avoid privacy conflicts, but they do not represent real privacy policies well.

Thomas et al. [147] study the risks of multi-party privacy in social networks. As a part of their research and as a way to show the dangers of unsuitable co-privacy management, the authors try to infer the information of Facebook users from two sources: links between friends and conversations with friends. The authors achieved an accuracy above 60% inferring information such as gender, political views or favorite TV shows. The authors propose a privacy framework to avoid these privacy conflicts. The framework is based on exposure policies instead of privacy policies. Each user referred to by a piece of information posted on any page of the SNS (e.g. Facebook wall) can define an exposure policy. For example, Alice posts on her Facebook wall a comment where Bob is referred to. Alice specifies a privacy policy because she is the owner of the information and Bob can define an exposure policy to limit the users that can access that comment where he appears. The exposure policies are defined in terms of the type of information and the page where it is posted. The authors prototyped their solution as a Facebook application that guarantees that every exposure policy is respected. The authors did not test their prototype, so experimental evaluation is lacking. Moreover, one of the main concerns of this proposal (also expressed by Thomas et al.) is that if several users are involved in the privacy management of an item, the group of users permitted to access that item tends towards the empty set.

Wishart et al. 2010 [158] propose a collaborative creation of privacy policies for shared items. The authors detect two roles, the owner of the item and the co-owners, which are designated by the owner and are individuals that are affected by the item. The main idea of the proposal is that owner and co-owners refine the privacy policy iteratively, and, hopefully, at the end, the preferences of everybody will be considered in the resulting policy. The authors define a model for privacy policies that contemplates the specification of strong and weak conditions. Weak conditions are overridden by strong conditions. In other words, weak and strong conditions establish a preference order, where strong conditions are those that a user considers an accessor must fulfill and weak conditions are those that can be overridden by strong conditions specified by another co-owner. As a proof of concept, the authors developed a tool. The tool has not been tested or evaluated with real users. The concept of strong and weak condition introduces preferences when defining privacy policies. However, the proposal does not consider other problems that come from co-authoring. As the owner and co-owners refine the privacy policy, many condition conflicts can arise, for example, two strong conditions that contradict each other. Moreover, the process of refining can be virtually infinite; it is possible that at the end only the preferences of the most persistent user are considered, since there is no method to prevent an endless modification of the privacy policy.

Besmer and Lipford [6] propose a method where the owner of a photograph that involves several individuals is in charge of managing its privacy and the other involved users can only suggest privacy preferences. The authors performed an experiment to collect information about privacy concerns of SNS users about photographs and what strategies they use to control the leak of private information. According to their study, SNS users consider that the owner of a photo (the individual who uploaded the image to the SNS) is the one in charge of assigning a privacy policy. Therefore, the other users who appear in the photo may only suggest privacy policies relying on the responsibility of the photo owner. The authors developed a Facebook application that allows a user to send privacy suggestions to the owner of a photo where that user appears. The authors also performed an experimental evaluation of their approach and their application. According to their findings, the participants were comfortable using this approach. The main issue with this approach is that the owner of the photo may not responsible enough or cannot deal with the petitions of other users and the possible preference conflicts that can arise from these petitions. As a future improvement of the software proposed by the authors, it should help owners to decide which policy to apply in the case of interest conflicts.

Hu and Ahn [69] propose a multiparty authorization framework that enables collaborative management of shared data. The authors divide the users into three groups: *owner*, the user that uploaded the item to the SNS; *accessors*, the users that want to access the item; and *stakeholders*, the users that are affected by the item somehow, for example, being tagged in a photograph. The users in any of these groups have to specify their privacy preferences and assign a sensitivity score to the item being shared. As several preference conflicts can arise, the owner of the photo is in charge of specifying a conflict resolution strategy. The authors propose several conflict resolution strategies (for example: owner-overrides, full-consensus-permit, strong-majority-permit, and many more). Since the owner decides the conflict resolution strategy, she has greater control over the resulting privacy policy associated to the item. For example, if the owner considers that the item is very sensitive regarding her privacy, then she will assign a restrictive conflict resolution strategy like *strong-majority-permit*. The authors developed a prototype as a proof of concept and tested the policy evaluation performance of the prototype. They conclude that the prototype is fast evaluating policies. However, they did not test their proposal with users and did not evaluate the satisfaction of the users with their approach.

Hu et al. [72] propose a multi party privacy management mechanism based on game theory. They use some of the concepts introduced in their previous work [69]. The mechanism is based on sensitivity and sharing loss. Every user connected to a piece of data specifies their perceived sensitivity and the allowed accessors. If there are conflicts in the allowed accessors (some users allow some accessors while others block those accessors), privacy loss and sharing loss are calculated. If privacy loss is higher than sharing loss, the access is denied, and vice versa. The more sensitive a picture is, the more potential privacy loss. The more potential accessors are blocked, the more sharing loss. This way of resolving conflicts can lead to problems when not every involved party is well-behaved. For example, a malicious party can increase the sensitivity of a photo to make the privacy-sharing tradeoff close to her expectations. To address this problem, the authors created two algorithms to ensure convergence to the Nash equilibrium. The authors evaluated their approach with a user study. The results show that the majority of participants prefer to cooperate and respect others' privacy preferences. However, the authors also detected that the participants did not adopt the best strategies when making decisions. This indicates that there is a gap between game theoretical approaches and real human behaviors.

2.5 Modeling Human Relationships on Social Network Services

Current SNS make little effort to differentiate between users. Users are either friends or strangers, with nothing in between. This approximation does not represent human relationships well. As introduced in the paper by Granovetter [56], the concept of *tie strength* defines the relationship between two individuals. In his work, Granovetter speaks about two different types of ties: *strong* and *weak*. On the one hand, strong ties usually include relations such as family and close friends. On the other hand, weak ties may refer, for example, to coworkers or less trusted friends.

As described in the previous sections, a rich relationship model can play a key role in privacy protection. SNSs aim at creating virtual versions of the real social networks of their users. An accurate representation of the real social graph of the users can help users to manage their privacy [68]. Moreover, ReBAC models need relationships to be modeled truthfully. Wiese et al. [156] studied the correlation between information sharing willingness and tie strength. Their research proved that, in certain situations, the strength of ties can be even more significant than grouping (the current approach of most SNSs) for predicting sharing. Based on their results, they suggest that a mixture of grouping and tie strength could allow richer and more usable sharing policies. For example, users could specify in a privacy policy for a given photo that only friends and close contacts (i.e., high tie strength) have access. Several approaches to create a social model that is based on the concept of the tie strength have been proposed. Most of the works on this matter try to infer a value for the strength of the relationships.

Figure 2.3 depicts a conceptual map that sums up all of the reviewed approaches that propose models for human relationships on social networks. As seen in the figure, the models are based on two different concepts: homophily and intensity of communication. On the one hand, homophily states thats that the more two person have in common (job, friends, hobbies) the more likely it is that these two persons

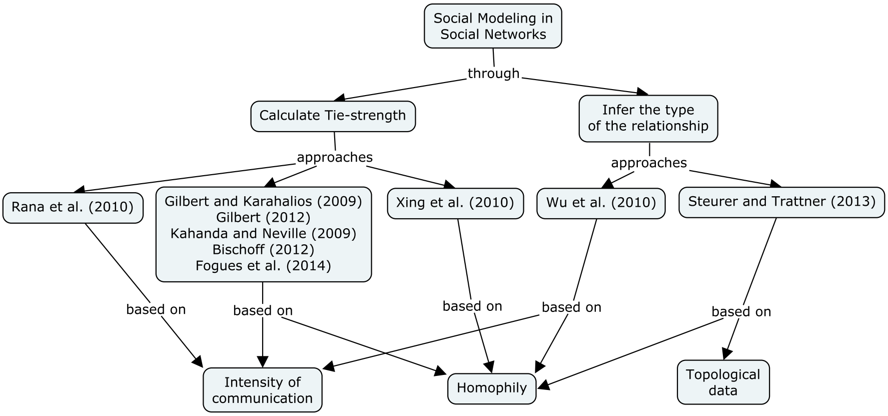


Figure 2.3: Concept Map of Human Relationship Models

have a strong relationship. On the other hand, the idea of communication intensity is that if two persons interact frequently, then their relationship should be strong. The majority of models use both concepts and combine them.

Gilbert and Karahalios [53] proposed a model that predicts tie strength among users of Facebook. The authors selected a group of variables available at Facebook. These variables covered different tie-strength dimensions; for instance, the number of links shared correspond to the “services given“ dimension. They evaluated their model with the participation of Facebook real users. The researchers asked the participants to assign a value of tie strength to their contacts. Besides, the researchers collected the values for the selected variables. Using the group of over 70 variables Gilbert and Karahalios achieved an accuracy of 84%. One of the limitations of this model is the huge amount of information it requires to predict tie strength. This high volume of information has to be processed and can become a bottleneck in the performance of the model.

Gilbert [52] expands his previous work [53] and proposes a model to infer the tie strength of relationships in a different SNS, Twitter. Gilbert selects a set of variables that are similar to the variables chosen in [53]. Some variables that were selected in Facebook do not exist in Twitter. In this situation, the author chose analog variables;

for example, the number of friends were replaced with the number of followers. In order to evaluate the new model, Gilbert developed a tool called *We Meddle* that predicted tie strength. Users were asked to try *We Meddle* and evaluate its predictions. This work showed that the model proposed in [53] can somehow be generalized and adapted to different SNSs.

In their work Kahanda and Neville [80] propose using transactional information to predict tie strength in social networks. The model is constructed using 50 variables. The variables are classified in 4 groups: (i) attribute-based Features, (ii) topological features, (iii) transactional features, and (iv) network-transactional features. The first three groups of variables are considered in other works [53, 161]; however, the fourth group, network-transactional features, represents a novel approach for tie strength prediction. The variables in this group capture the transaction information between nodes, but they represent it within the context of the larger network structure. For example, one of the variables in this group is the interaction among all the nodes in the network, not only between two pairs. According to the results of the Kahanda and Neville study, the most predictive variables are those in the fourth group. The study lacks an evaluation with humans; therefore, the accuracy of the results may have been affected as the tie strength is a purely human-dependent concept.

Xiang et al. [161] proposed a model to infer relationship strength based on profile similarity, with the goal of automatically distinguishing strong relationships from weak ones. The model proposed by Xiang et al. uses the concept of homophily to infer the tie strength between two individuals. Xiang et al. test their model with proprietary data of the SNS LinkedIn and data from students of Purdue University on Facebook. With the LinkedIn data they tested their tie strength prediction against other heuristics. The authors considered the number of times a user checked another user's profile page as the indicator of the tie strength between them. For the evaluation in Facebook, they calculated the ground truth tie strength between two users as a combination of the number of common networks, common groups, and common friends for those two users. The results of the evaluation show accurate results.

However, the evaluation was synthetic; the ground truth tie strength values were calculated using heuristics and were not specified by the users themselves. Therefore, it lacked the corroboration of the results from the participants. The main difficulty of this model to work accurately is that it relies on profile information, which tends to be incomplete or of low quality. Few users specify their address, job, college, or other variables needed by the model to work [39].

Rana et al. [119] propose a theoretical framework for calculating social strength and a ranking of contacts sorted by their social strength. The algorithm proposed copes with the complexity of users using a wide variety of communication services. For example, a user can contact her friends using Twitter, Facebook, and SMS. Moreover, the algorithm also considers different ways of communication inside the same communication service. For example, in Facebook, a user can contact another sending a private message or through a comment anout a photo. These two ways of establishing contact are considered to be different *communication tools*. The algorithm counts the number of interactions on each communication service and assigns a level of importance to the communications established on that service according to its ratio of usage. Finally, a value of tie strength for each contact is obtained adding the number of interactions with that contact and weighting the interactions according to the service and tool used. The main limitation of this algorithm is that it only considers the frequency of interaction between the user and each contact. However, as shown by other studies [85], a high number of interactions does not necessarily imply a strong tie.

Bischoff [9] analyzes online friendship in the Lastfm musical social network. Lastfm offers social features like friendship links, message exchange, and a personal profile. Moreover, Lastfm allows users to specify their musical tastes, favorite artists, and music event attendance. The author collected public information about 48,527 Lastfm users. The information collected contained variables such as: friend links, messages sent, tags assigned to artists, music recently heard, preferred albums, preferred artists, events attended, and demographic data (gender and country). The author used the gathered data to classify each pair of users as: no link between them, weak relationship, or strong relationship. The number of events coattended by both persons determines the strength of the relationship in the training set. For example, if two persons attended 2 events together, then they have a weak link; however, if they attended 11 or more, then they have a strong link. The author also used the data to create a friendship recommender system. In both experiments the results were promising. The most predictive variables in Bischoff's study were those related to homophily, such as coincidences in the preferred artists or same country. This study shows that, depending on the objective of the social network, different variables have to be considered for the task of inferring the tie strength.

Fogues et al. [41] introduce a tool called Best Friend Forever (BFF) that automatically groups and assigns a tie strength value for the contacts of a user. In order to infer tie strength values, BFF follows an approach similar to [53] and [80]. However, BFF uses a much smaller set variables, only 11. The reduction in the number of variables makes the variable collection task faster and less costly, thus increasing the utility of the tool. For automatic group creation, Fogus et al. used the algorithm proposed by [126]. This hierarchical diffusion algorithm is founded on the triadic closure principle, which suggests that, in a social network, there is an increased likelihood that two people will become friends if they have friends in common. The authors made an experimental evaluation of the tool and compared the results obtained by their tool with the preferences of 17 participants. Despite the reduction of variables for tie strength prediction, the tool performed accurately. On the matter of groups, the tool also worked with precision. Fogu'es et al. present a relationship model that focuses on a specific SNS, Facebook. However, although several relationship defining variables used for Facebook can also be found in other SNSs, a more general model that works in different social networks is needed. Moreover, the authors mention a correlation between tie strength and community creation, but they did not study this fact.

As explained in the previous section, some ReBAC models allow the specification of multiple types of relationships. Wu et al. [160] deal with the task of differentiating between personal and professional closeness. The authors performed a survey with users of a professional SNS called Beehive. This SNS was deployed at the IBM company in 2007. The authors asked the participants to assign a value of tie strength to their contacts in the social network from a professional and a friendship point of view. The results and the most predictive variables identified by the authors are close to the ones obtained by Gilbert and Karahalios [53]. To differentiate a professional relationship from a personal one, the authors identified a set of variables that worked as strong predictors of a professional relationship. This research shows that different types of relationships may need specific predictors, making the relationship classification a complex problem.

Steurer and Trattner [137] study what topological variables can be used to differentiate between acquaintances and partners. Since collecting topological data from facebook-like social networks is nearly impossible, the authors employed the virtual world of Second Life. Second Life has virtual locations and every time a user visits one of these locations the information is stored and can be collected. After crawling data from Second Life, the authors created two social graphs: one based on topological information, and the other based on homophily information. Users on Second Life can get married, the authors employed this information to train a number of classifiers to differentiate between acquaintances and partners using the collected data. Their results show that topological information can be used to infer the type of the relationship with a high accuracy. With the ubiquitousness of mobile devices that can sense the location of their users, topological information is more available. Thus, future research should be able to use topological data obtained from real world.

This section reviewed studies that aim to deal with an essential ReBAC requisite, the modeling and definition of relationships. In Section 2.8, we identify future research paths that can improve the way relationships are modeled and classified. For example, considering how one person discloses information with another can help to define the type of relationship that exists between them.

A ReBAC model cannot exist without a suitable social model. However, a model that truthfully represents human relationships is still useless if its users cannot manipulate and understand it with ease. As explained in Section 2.1.2, another requirement for a ReBAC is usability. Sections 2.6 and 2.7 show studies that aim to increase the usability of privacy mechanisms for SNSs. The studies reviewed propose tools that help users to configure, adapt, and understand their privacy preferences in SNSs.

2.6 Recommender Systems

A first troublesome task that SNS users must face when dealing with their privacy preference configuration is the definition of privacy policies. As shown in section 2.2, a ReBAC model can consider a large number of variables and use a complex privacy policy definition language (e.g. a language based on regular expressions or ontologies). It is not realistic to assume that an average SNS user is familiar with these concepts and can effectively use them. A common approach to help users in this task is to suggest privacy policies to them or guide them during the process. Users should have the last call on what can be disclosed and what is private. While recommenders can help reduce the user's burden, they are rarely perfectly accurate. Thus, it is important for the user to be able to view, understand, and modify the recommended policy before it is applied; it is also important for the user to be able to maintain the policy over time.

Figure 2.4 depicts a conceptual map for all of the studied approaches that propose a privacy recommender system for SNSs. The following subsections review the studies according to the classification shown in the conceptual map.

2.6.1 Based on What Others Do

A first approach for privacy policy recommenders is based on how other users set their privacy policies. For example, if every friend of a user decides to hide their

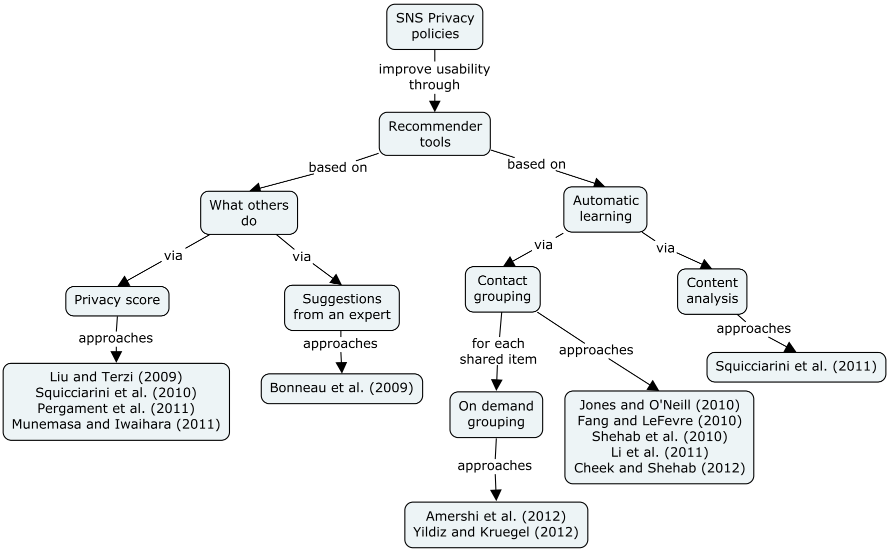


Figure 2.4: Conceptual Map of Privacy Recommender Systems

address information, then it is likely that a good recommendation for the user is to also hide her address information. In this section, we review studies that use this approach.

Bonneau et al. [11] propose a tool that suggests privacy policies based on expert users' configuration. Users can specify their privacy policies and then share them over the SNS. Any user can apply the privacy policies shared by another user, rate them, and recommend them to her friends. It can be expected that the best privacy policies and the experts who created them will have high rates and will be used by a high number of users. Moreover, users can subscribe to their favorite privacy expert. In this way, users' privacy policies will automatically update when their preferred expert updates her policies. This approximation only allows the automation of privacies with low granularity as only general policies can be shared by experts.

Squicciarini et al. [135] present a privacy manager named PriMa that suggests privacy policies taking into account the sensitivity of content according to what users tend to do on the SNS and tie strength. To use PriMa, first, the user expresses her concerns about disclosing each of the attributes in her profile. If an attribute is left without a sensitivity value, PriMa infers this value from the values assigned to that trait by the contacts of the user. Once each trait has a sensitivity value, they are grouped in clusters depending on their sensitivity. In order to recommend policies, PriMa computes a user access score, which is a representation of the adequacy of a given target user to access a given cluster of attributes of the main user's profile. The score is based on the type of the relationship between the target user and the main user. The type of the relationship has an associated tie strength value that is predefined by the SNS provider. Finally, if the user access score is higher than a threshold, then that user has access to that attribute. The paper lacks an experimental evaluation also, the assumption of predefined types of relationships and tie strength values can lead to policy generation errors. For example, it is not possible to assume that every user agrees that a family relationship tie has a strength of 0.8 and a friend tie has a strength of 0.5.

Munemasa and Iwaihara [104] follow the line of Bonneau and Liu research [11]. These authors propose software that tells the user if her privacy settings are introvert or extrovert by computing a privacy score and comparing it with other SNS users' scores. The software created by the authors collects privacy settings of several SNS users and rates them. The rating of a privacy setting is based on the volume of information disclosed and how likely it is that others disclose the same information. The authors base the calculus of the privacy rating on the work of [98]. When a users configures her privacy settings, the software rates her configuration and compares it to the settings previously collected. From this information the software can determine whether the user's privacy setting discloses more or less information than the average user of the SNS. The authors performed an evaluation of their tool with 15 Facebook users. The evaluation participants only answered a questionnaire about the suitability of the tool. The article lacks experimentation that evaluates how accurate the recommendations made by the tool are (e.g. comparing them to the actual preferences of privacy concerned participants).

Pergament et al. [115] present a system named FORPS that helps user to decide what can be accessed by a friend depending on the behavior of that friend on the SNS. FORPS calculates a privacy score for a target user for different themes. For example, a user can be very discreet with regard to religion and very indiscreet regarding political views. To calculate the score for each different theme, FORPS analyzes whether the target user discloses information on that matter with main the user's friends, her number of commentaries about that theme, and the sentiments on those commentaries. The system also analyzes the behavior of the friends of the main user in regard to the target user. Finally, FORPS recommends to the principal user a level of privacy for each theme towards the target user. The proposal lacks an experimental evaluation. Therefore, the utility of the system is not evaluated. Additionally, the system only recommends a numeric level of privacy for each theme. It is difficult to map a numeric value for what things is recommended to disclose and what are not. For example, on a 0 - 1 scale, what does a 0.3 degree of disclosure in politics mean? Should I disclose my preferred political party or not?

2.6.2 Based on Automatic Learning

Another approximation for recommenders is based on learning. Proposals in this section observe how the users interact with others and learn from their preferences. Once the recommender has learned enough, it is capable of automatically recommending privacy preferences. This section reviews studies that use the concept of automatic learning.

Jones and O'Neill [77] investigate what criteria humans use when they divide their social network in groups so they can share different information and avoid embarrassing situations with each group. The authors recruited 15 participants and asked them what factors they consider when creating groups in Facebook. According to their findings, the factors that affect grouping sorted from more important to less are: cliques (densely connected groups), tie strength, geographical location, organizational boundaries, temporal episodes, and functional roles. The authors also tested the clustering algorithm SCAN and compared its output with the groups manually created by the participants and they achieved 44.8% of similarity. In order to improve the accuracy of the algorithm the authors used some information from their interviews and added some information about tie strength to the algorithm. They could not add tie strength information for all participant data. With tie strength, the algorithm achieved an average similarity of 67%. This study shows that several factors are taken into account by users when creating groups. However, the authors did not collect enough data from participants' profiles and could not improve the clustering algorithm by modifying it, so it considers all the discovered factors.

Fang and LeFevre [39] propose a *wizard* software that suggests users privacy policies for different items on their profiles, like birthday, address, or telephone number. Since the proposal of Fang et al. uses supervised learning, participation of the user is needed. First, users' contacts are hierarchically grouped. To form the groups, the wizard considers community, profile, and activity features. For example, mutual friends is a community feature, while hobbies or fan pages that a user likes are considered to be activity features. Once the groups are created, the wizard asks the user to assign access grants to some of their contacts. The main idea of this process is that the user assigns access grants to the more representative users of the groups previously created. In this way, the user only needs to assign a low number of grants, and the process is much faster. According to the results of the proposal's evaluation, the wizard behaves better when only community features are considered. This can happen because many users do not specify their hobbies or demographic information. The privacy wizard is designed to protect only user's traits, like birth date, address, and telephone number. Other elements like images or videos are not considered by the wizard. Moreover, to manage items like photos or videos, the wizard would need to be enhanced to consider specific features of the item, for example, tags on a photo.

Shehab et al. [125] introduce a privacy policy recommender system that is based on supervised learning. Their system works in 5 steps. In steps 1 and 2, the attributes of the main user's contacts are collected. Then, they are clustered according to their attribute similarity and a representative user is selected for each cluster. In step 3, the main user assigns access rights to the representative users. These labeled users are utilized by the classifier as the training set. In step 4, the rest of the contacts that are not labeled are classified and labeled accordingly. Finally, in the step 5, the main user's classifier look at the classifiers of the main user's neighbors and fuses itself with those that have similar clusters of users. One threat to privacy of using this recommender system is that it needs to access the privacy preferences of other users and these preferences should be also private.

Squicciarini et al. [134] propose a system called A3P that predicts a privacy policy for images in the context of social networks. A3P takes into account two variables when recommending a privacy policy for an image: social contexts and image content. A3P analyzes the content of the images and assigns a category to them. For the context, the authors predefine a set of social contexts (e.g. family, coworkers) and assign an intimacy value to each one of these contexts. Since A3P is based on supervised learning, it needs the user to specify some privacy policies before starting to predict policies. The policies can specify what contexts are allowed to access the image and what privileges each context has. Once A3P has learned enough from the user, it starts predicting policies using a policy mining algorithm. The policy mining algorithm considers the tendency in policy strictness of the user. Therefore, if the user tends to disclose more information, the suggested policies will be more extrovert and vice versa. Squicciarini et al. evaluated their proposal with humans; however, the photographs provided during the experimental evaluation were previously selected by the authors and do not correspond to real photos of the participants. An evaluation of how A3P performs with photographs taken by the experimental evaluation participants should be an extension of this paper.

Li et al. [90] present a privacy policy recommender based on semantics. Their approach is similar to the wizard presented in [39]. The main difference between the two approaches is that the one from Li uses semantics to find similarities among users. For example, a user can specify that he likes basketball and another that she likes NBA. Both hobbies are similar, but an approach without semantic knowledge will overlook this similarity. The authors present a *k*-Nearest Neighbors algorithm that uses semantic knowledge to compute the distance between two persons. When the user wants to specify access permissions for a new friend, the recommender suggests a policy that is already defined for the semantically closest *k* friends. The paper presents an experimental evaluation with 76 Facebook users and compares the performance of this approach against others. An interesting future improvement of this work would be the study how the use of techniques like uncertainty sampling can reduce the number of contacts to label so the recommender can accurately suggest privacy policies, thus reducing the effort required from the user.

Cheek and Shehab [24] introduce a privacy policy recommender that uses the similarity among friends to ease the process. Their approach is based on a graphical interface and offers two functionalities: an assistant to create groups of friends and *same-as* policy management. The group assistant guides users in the burdensome task of labeling their friends. The assistant presents a set of ten predefined social groups: Family, Close Friends, Graduate School, Under Graduate School, High School, Work, I do not know, Friends of Friend, Community, and Other. Each contact is presented to the user and the user is asked to select a group for that contact. In order to speed up the process of labeling every friend, the assistant uses the Clasuet Newman Moore clustering algorithm to recommend groups for contacts that have not been labeled yet. Once every contact is labeled and belongs to a group, the *same-as* policy management asks the user to select a representative contact of each group and assign a privacy policy for that contact. The rest of the contacts on each group will be assigned the same privacy policy that was assigned to the representative contact of their group. The authors performed an extensive experimental evaluation where the users were asked to specify privacy policies for predefined groups of items: every album, demographic data, and educational data. The lack of granularity in privacy policies and the use of predefined groups of personal data limit the validity of the evaluation.

Amershi et al. [2] present a machine learning system called ReGroup that creates on-demand groups in social networks. The main difference between this proposal and others is that groups are not created once and used many times; instead, a new group is built for each item that is going to be shared. ReGroup uses a Naïve Bayes classifier to find similarities between users and make recommendations. When the user uploads an item to the SNS, ReGroup asks her to choose a contact to grant access to that item. After the first contact is chosen, the other contacts are sorted by similarity to the first one. Each time a similar contact is not selected, ReGroup adds a penalty to the similarity of that contact. In this way, this contact will appear in a later position the next time. The authors tested ReGroup and the standard application of Facebook, where users are sorted alphabetically in terms of group creation time and happiness of the user with the group created. The results obtained and the opinions of the participants highlight that each option works better depending on the size of the group. Alphabetical order is better for small groups, and ReGroup is better for large groups. This proposal alleviates the process of assigning a privacy policy for items; however, the user still has to select each user that is allowed to access the item separately. A future expansion of the article could be to compare what option is faster and easier for the user: correct a privacy policy suggestion or create a privacy policy from suggestions.

Yildiz and Kruegel [164] introduce a new algorithm that creates groups considering the users referenced by the item being shared and their social connections. The idea is that the groups created can help users to decide the privacy policy for the uploaded item. When a new item is uploaded, the algorithm creates a list of *participants*, who is the owner of the item, every user involved in the item (for example, users tagged in a photo), and a list of *candidates*, who are the friends of the participants. In each iteration of the algorithm, a candidate is inserted to the list of participants (users that are allowed to view the item). The inserted candidate is the one who maximizes a heuristic function. This function returns a high value when the analyzed candidate has many common friends with the users in the list of participants. Besides, the algorithm considers the tightness of the list of participants. If the participants are tightly connected, added candidates also have to be tightly connected to the participants; otherwise, the algorithm will look for loosely connected candidates. The algorithm keeps adding candidates until it cannot find an appropriate one. The authors performed an evaluation comparing the results of their algorithm with other local community finding algorithms. The authors considered for the best result for an algorithm to recover the entire group of the participants of the item while keeping to the minimum the number of members that do not belong to the participant group. In other words, they considered that the social cliques were entirely defined by the participants of the items. One problem with this assumption is that the privacy policies will probably tend to be very restrictive. Moreover, the article lacks an evaluation of the quality of the proposed groups made by human participants.

In Section 2.8, we have identified future lines of research for privacy policy recommenders. A crucial challenge is to combine privacy recommender tools with ReBAC models. As ReBAC models are key to the suitable management of information disclosure on SNSs, privacy policy recommenders should create policies according to the policy language defined by the access model.

The papers analyzed in this section focus on making privacy policies easy to configure. However, privacy policies should not only be manageable, they also have to be easy to maintain over time so that the users can adapt them as new relationships are established and previous ones change. Moreover, as the users set and refine their privacy policies, the global view of how their information is accessed by others grows in complexity. Ideally, users have to be perfectly aware of who has access to their private information and to what specific parts of that information. Section 2.7 expands the concept of usability by introducing the notion of privacy setting understanding and also reviews proposals that help users to create better mental models of their privacy settings in SNSs.

2.7 Improving Privacy Settings Understanding

A great obstacle that users find when dealing with privacy on SNSs is the difficulty of figuring out how and what personal information is disclosed over the SNS. SNSs have improved their user interfaces; for example, Google+ offers the Social Circles graphical tool that facilitates the laborious task of grouping contacts. These new interfaces facilitate the process of managing privacy settings and figuring out what others can see from our profile. However, there is still room for improvement, and if in the long term SNS developers aim to implement more complex access control models, they have to accompany them with easy-to-understand and easy-to-use interfaces. Currently, we find three different approaches to make privacy policies more understandable: graphical interfaces, privacy policy simplification, and privacy setting misconfiguration detection. Graphical interfaces use visual tools to enhance the understanding of the privacy policies. Privacy policy simplification tries to remove unnecessary complexity of privacy policies and make them more legible for the user. Finally, privacy setting misconfiguration detection tries to infer when users make a mistake while configuring their privacy settings to warn them and encourage them to correct the misconfiguration.

Figure 2.5 depicts a conceptual map for all of the reviewed approaches that propose tools or mechanisms that help users to understand their privacy settings in SNSs. The following subsections explain the approaches according to the order shown in the conceptual map.

2.7.1 Graphical Visualization

Lipford et al. [94] propose a new privacy management interface for Facebook called AudienceView. This new interface offers users the possibility to observe how their profiles are seen from different points of view. Each different point of view corresponds to one of the possible audiences: search, network, an individual friend, and friend group. To evaluate their proposal the authors recruited 16 participants and

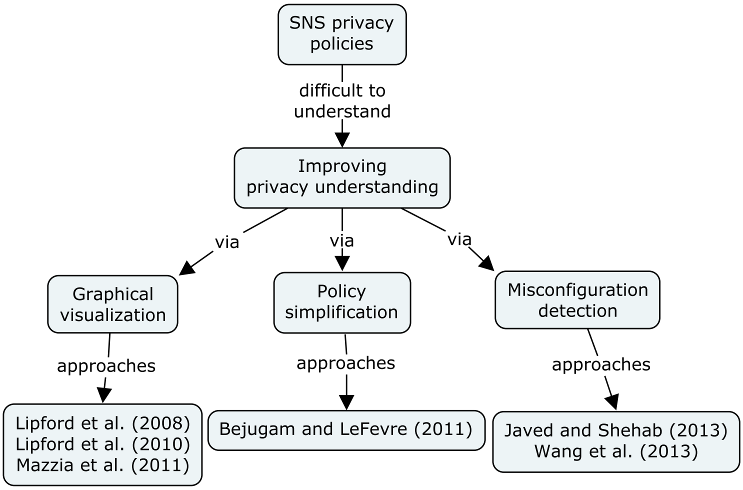


Figure 2.5: Conceptual Map for Improving Privacy Setting Understanding

measured their performance in different tasks using the standard Facebook interface and the proposed new one. The results showed that users felt more comfortable with the new interface and it helped them to understand better the actual consequences of their privacy settings. Currently, the prototype is limited as it only shows how the profile data (name, hometown, hobbies...) is seen from different audience points of view. It would be very useful to expand the prototype so profile publications like photos or comments could be inspected by AudienceView. This improvement would probably require a new interface and a clever graphic design to make it usable and understandable for the user.

Lipford et al. [95] compare their previous proposal [94] with another kind of privacy policy representation, Expandable Grids. Expandable Grids show precisely what a policy allows or does not allow in a matrix using hierarchical axes that can be expanded or contracted to show more or less policy detail. Expandable Grids were initially proposed for access control policies in file systems. Therefore, a more extensive review of this representation is out of the scope of this paper. For further information of Expandable Grids, please refer to [120]. The experimental results obtained by Lipford et al. showed that both representations were highly usable, and they did not find clear advantages of one over the other. However, experiment participants did have clear and different preferences; some of them preferred the compact view of Expandable Grids and others the visual feedback of AudienceView.

Mazzia et al. [103] present PViz, an interface that is focused on helping users understand the visibility of their profiles. PViz presents the social structure of user's contacts in a graphical way. Contacts are automatically grouped in communities using the idea of modularity optimization. This methodology generates a hierarchical structure of groups. The higher levels of the hierarchy have groups of users that are loosely connected and share less attributes, the lower levels of the structure represent groups of users that share many common friends, tastes, and demographic data. This approach for automatically creating communities is also used in [39]. Moreover, PViz automatically labels each group, selecting the most common attribute of the contacts that conform the group. For example, if the majority of contacts in a group are from Washington, that group will be labeled as Washington. The graphical interface allows the user to select an attribute and see what communities are allowed to view that attribute. Communities are colored; the darker the color is, the larger the number of members of the community that can view that attribute. As the communities are hierarchically organized, the graphical interface allows users to zoom in or out of a group to see more or less detail of that group. The authors empirically evaluated PViz comparing it with AudienceView [94] and the Facebook standard interface. Their results showed that participants preferred PViz over the other two options; some participants even suggested that a combination of AudienceView and PViz could create a better solution.

2.7.2 Privacy Policy Simplification

Bejugam and LeFevre [5] present a policy simplification utility. Current privacy policy specification tools are based on rules that allow the user to specify positive rules (”Show this to”) and negative rules (”Hide this to”). These specifications can lead to long and complex policies, and even include redundancy, making them difficult to understand and adapt when new contacts or groups are created. The approach of Bejugam and LeFevre is based on two functionalities: automatic community creation and a policy consolidator. The automatic community creation algorithm is the same algorithm used in [103] and [39]. The policy consolidator transforms a privacy policy specification and produces the smallest equivalent privacy policy. The authors performed an experimental evaluation with nine subjects. Their test proved the utility of their approach and the hypothesis, which states that users will comprehend, remember, and maintain simplified policies easier than their verbose counterparts.

2.7.3 Privacy Setting Misconfiguration Detection

Javed and Shehab [75] introduce a utility that helps users avoid employing incorrect privacy policies. The utility is divided in two elements. The first is a policy composition tool that employs *tag clouds*. The tags are the name of the groups that the users have in their networks (e.g., friends, colleagues, and so on). The size of a tag depends on the number of times it has been used in other privacy policies and the privacy risk it represents. For example, if a user tends to share his photos only with his friends, the tag friends will be much bigger than the tag public. The second element of the utility is a misconfiguration scanner. The authors defined a set of seven misconfiguration patterns. For example, there are common friends in the groups allowed to access a photo, which might have been the intended audience. Once the user specifies his privacy policies, they are scanned. If any of the patterns is found, the user is warned. The authors evaluated their approach with a study. Their results show that the misconfiguration scanner helped users to notice that certain policies allowed access to an audience that was not their intended one. However, the decrease in the number of misconfigurations using the composition tool was not significant.

Wang et al. [151] propose a paternalist approach to help SNS users make better information disclosure decisions. The authors designed three privacy nudges for Facebook: picture, timer, and sentiment. The privacy nudges are limited to status updates. Picture nudge is similar to the surveyed graphical visualization approaches. When the user is going to update his status, this nudge shows him a random selected subset of the people that will be able to see that content with the current privacy configuration. Timer nudge encourages users to reflect on status update they are about post and on its privacy configuration. The nudge gives 10 seconds to the user before the status update is actually posted. During these 10 seconds users can modify the status or cancel it altogether. Finally, the sentiment nudge is based on a sentiment-analysis module that analyzes each new post. Each post gets a sentiment score and the user is warned about how other people may perceive it. For example, a status update that says “I am angry” will get the sentiment nudge “People may perceive your post as Negative”. The authors evaluated the nudges through a study and subsequent interviews. The majority of participants found that picture and timer nudges were useful and could change positively sharing behavior. On the other hand, the majority of participants disliked the sentiment nudge and did not find it useful.

2.8 Open Challenges

Although the proposals shown in this survey cover some of the requisites that an access control for SNS should fulfill, we have identified many possible lines for future research. In this section, we outline some of the most challenging possible future directions in the research field of access control models for social networks and their usability. These possible paths of research are open challenges that were identified during the realization of this survey. The accomplishment of these open challenges will have a great impact in determining SNSs to be useful, entertaining, and privacy safe services available at the Web 2.0. On the one hand, SNS users will feel safer in the context of SNSs and will feel more inclined to register and use these services. On the other hand, we believe that the utility of SNSs will increase since users will upload and share much more information in their profiles.

2.8.1 ReBAC and Content Type

A number of works show that users take into account the type of the content when they are defining how that content will be disclosed [61, 83, 163]. However, current access control do not consider content type. On the one hand, including a content type as a new attribute of access control can improve the flexibility and expressiveness of privacy policies. On the other hand, more attributes can increase the complexity of access controls and privacy policies. Therefore, future research should evaluate the effect that the inclusion of the attribute content type can have over access controls. Specifically, new research should evaluate the complexity of the privacy policies created with the new access control, the required number of privacy policies that users need to express their privacy policies, the number of privacy conflicts generated by the policies, and the understandability of these policies.

2.8.2 Inferring Tie Strength from Different Sources

The studies about tie strength in SNSs use information available on the SNS; for instance, they consider the number of messages exchanged between users, the number of photographs shared or the number of common friends. A common pitfall of these works is that relationships that mainly occur outside the SNS are not well evaluated [53]. A possible solution for this problem is to search for information outside the SNS. Recent research [102, 34] infer tie strength values from data available on the Internet or from real life. Mixing data from the SNS and from other sources would create approaches that infer tie strength more accurately. However, the use of external data can create some privacy concerns since it directly relates to re-identification and profiling privacy breaches (these breaches are defined in section 2.1.1).

The proposal of Murukannaiah and Singh [105] is a first attempt to use real world information to infer the social data of the user. The authors present a tool called Platys Social that runs on a mobile device. This software learns a user's social circles and the priority of the user's social connections from daily interactions. The software infers the interactions from information that is available on mobile devices, such as wi-fi networks, bluetooth connections, phone calls, and text messages. Combining all sources of information (real world through a mobile device, Internet, and social network) could positively improve the tie strength inference and classification of relationships.

2.8.3 Adaptive Relationship Models

In human relationships, the disclosure of private information represents an important part of these relationships [57]. One of the main reasons why people exchange private information is because they perceive that in the future this information disclosure may become a gain in social capital. The way users share with others change the perception of both parts about the way they should interact in the future. For example, if user *A* constantly discloses information to user *B*, but user *B* only reveals a small portion of his information to user *A*, it can be assumed that in the mid/long term user *A* will reduce the amount of information disclosed to *B*, or even stop communication at all.

Such et al. [145] propose a self-disclosure decision-making mechanism for multiagent systems. The proposal is based on psychological findings regarding how humans disclose personal information in the building of their relationships. The implementation of this mechanism (or a similar one) to a relationship model would increase the accuracy of that model, especially over time. The decision of specifying a privacy policy would consider not only the current situation of the user's relationships, but also how the new specified policy itself would affect the relationships and the gain or loss of tie strength for those relationships.

2.8.4 Tie Strength Utility

The studies reviewed in section 2.5 infer the tie between two individuals in only one dimension, the strength. However, humans behave differently with their contacts depending not only on the strength of the relationship but also on its utility. Ties can be viewed not only for their strength but also how useful a tie is depending on the situation or the need. Imlawi et al. [74] studied how instructors can increase the engagement of students on course-based online social networks. Their research showed that students feel more engaged when the instructors make humorous posts. However, this increase in engagement is not seen when the instructor posts private information unrelated to the course.

A first attempt in this direction is the paper by Rosen and Chu [122]. The authors claim that the strong-weak tie dichotomy is conceptually misleading, and propose a multi-dimensional taxonomy of social network ties. The authors propose that a tie should not be evaluated by its strength but by its utility. In this sense, the authors study the possible social dimensions that matter during specific contexts. Future research should address to what degree each dimension affects the tie utility as well as how to identify the context an interaction belongs to.

2.8.5 Self-presentation Management

Kairam et al. [81] detected that the most common motivation to use SNS and share information is to create a self-presentation. Users care about what image they project on others and how they affect others. Klout[6], which is a tool that helps users to see how they influence others in popular social networks, is a first step towards a self-presentation management system.

The studies reviewed in 2.7 aid SNS users to understand what other users can see in their profiles according to their privacy preferences. A similar approximation could

[6] http://klout.com/home

be used so users could express their privacy policies based on facets. Users would find it more natural to specify which facet of their life they are willing to show to each contact, instead of a sequence of rules that specify who can see what and who cannot. For example, a user could choose to show a funny facet to her family members or friends, allowing them to see comical photos or posts, while showing a professional facet to colleagues or potential employers.

2.8.6 ReBAC: Usability and Visibility

A change in the access controls of SNSs will represent a change in their privacy controls as well. ReBAC models, such as the ones proposed by [23] or [25] use technologies like semantic web or regular expressions. Suitable privacy controls for SNSs should hide the complexity of such technologies and facilitate their use. Moreover, the inclusion of new attributes to the access control (tie strength and content type) also suppose a modification of privacy controls. Further studies should address how usable privacy controls can be created for new ReBAC controls for SNSs.

In the same fashion, privacy visualization tools will have to adapt to ReBAC models. For example, visualization tools should explain to the users in an understandable way how their information is disseminated according to a specific type of relationship. Moreover, complex models that allow the specification of a privacy policy hierarchy, user-to-resource relationship, or social paths will require visualization tools with clever designs. Google+ and the friend circles application is a good example of how privacy visualization tools can be designed to be usable and engaging for users.

2.8.7 ReBAC and Co-privacy

As stated in Section 2.4, co-privacy can cause several privacy conflicts among the users involved. SNSs should offer integrated solutions for situations where a shared item can cause tension among users. This tension and the complete lack of control of these situations can lead to users adopting strategies like un-friending the user who uploaded the controversial item or deleting their profile from the SNS. Hence, it is necessary to add shared item or co-privacy management to the the access control models for SNSs. However, there is no proposal that brings together a formal ReBAC model and shared content privacy management.

Some of the proposals reviewed in section 2.4 could be merged with a formal ReBAC. Users should be able to specify how strict their privacy preferences are on a shared item. Moreover, the ReBAC should have a privacy conflict resolution that guarantees that the resulting privacy policy maximizes the utility for every used involved. For example, a ReBAC policy language could allow users to specify a value that indicates how strong or weak their conditions are for each item or with respect to a certain type of relationship.

2.8.8 Privacy Settings Interoperability

There are many different SNSs, and they accomplish different objectives; for example, Facebook and Google+ are focused on maintaining friendships; LinkedIn[7] is focused on professional life; Meetic[8] aims to help users to find dates with similar people; and Flickr[9] is a social network that is centered on photograph sharing. Besides, the wider use of mobile devices that work together with SNSs also increases the variety and diversity of SNSs and their uses.

It is a common practice for SNS users to have profiles on different SNSs and to use each one for different purposes according to the general objective of the SNS. As pointed out in this paper, setting privacy settings is a burdensome task and users struggle doing it. Therefore, if a user has to configure her privacy settings separately for each one of the SNS she is using, this task becomes even more tiresome. Users need privacy settings that are interoperable among every SNS that they are members

[7] http://www.linkedin.com/

[8] http://www.meetic.com

[9] http://www.flickr.com/

of. A first step in achieving interoperability for privacy settings could be that access control models of SNS use a universal and well-defined privacy ontology for SNSs. This ontology should be able to differentiate among the different contexts that each SNS belongs to. To improve the interoperability of privacy settings, personal privacy agents could be developed. These agents will take care of their principals' privacy, adapting their behavior to the context their principal is at that moment.

2.8.9 Sticky Policies

Private information stored in the profile of a SNS user can move or be moved to different contexts. In order to maintain the privacy of that information, it is necessary for the privacy policies associated to the information move along with it. Policies of this kind are known as sticky policies and they enable users to improve control over their personal information as it moves across multiple parties. Sticky policies are orthogonal to ReBAC models, however, they become more necessary as mobile and pervasive systems gain in popularity. Users can share their information using a wide variety of devices and applications. Sticky policies can guarantee self-disclosure consistency among every device and application.

A common situation (in the context of SNSs) where personal information moves to a different party is the use of applications created by third parties and integrated in the SNSs. In [40] the authors propose a method to keep the maximum amount of personal information hidden to third party applications in Facebook. However, to the best of our knowledge there is no study that proposes how information is managed after the third party acquires it. The study of Pearson and Mont [112] presents a general framework for using sticky policies called EnCoRe. A similar approach could be used for SNSs, where the SNS itself would work as a trusted authority in charge of controlling that the information is disclosed by third party applications according to the owner's preferences and also of regulating the access to such data.

2.8.10 Personalization

As explained above, access controls need a number of attributes to capture correctly social relationships. Moreover, policy languages have to be flexible enough to enable users to express their actual privacy preferences. All this flexibility and diversity in privacy controls is needed. However, it can also increase the complexity of the configuration of privacy setting to a level that is unmanageable for some users. Personalization can alleviate this problem.

Users should be able to choose what elements from those offered by the access control they want to employ. For example, a given user may decide that tie strength is of no use for him when creating privacy policies. Thus, the interface for creating privacy policies should hide tie strength. To further ease the process of personalization, this could be done through automatic learning.

2.9 Conclusions

With the constant increase in the use of SNSs, mobile devices that connect us with others at all times, and in general, the spread of pervasive computing, we consider that privacy will be of paramount importance during the next few years. In a connected world, controlling our privacy and what others are able to see about us will be more difficult and complex. Hence, powerful, easy-to-use, and intuitive privacy solutions will be the subject of many research efforts during next years.

In this paper, we have reviewed approaches that offer partial solutions to the most critical problems of privacy management on SNSs. However, current SNSs have not adopted them and still lack the suitable privacy management tools. Approaches like Google+, where the control of information dissemination has been given great visibility [99], are first steps towards SNSs that are more respectful of privacy. In the not-so-distant future we envision a SNS that offers a privacy mechanism that satisfies every requisites mentioned in this paper and provides the features that users demand. In order to develop this ideal SNS, developers and researchers will have to deal with several challenges. The inclusion of ReBAC models in popular SNSs will improve the control of privacy for the users. However, ReBAC models are complex and SNSs will require a thorough design that guarantees the usability of the privacy management. Moreover, sticky policies and privacy settings interoperability will represent a technological challenge. We believe that universal privacy policy languages and access control models will be required to ensure these two requisites. The model and the language proposed would need to be flexible enough to allow different SNSs that focus on different social aspects.

CHAPTER

3

BFF: A Tool for Eliciting Tie Strength and User Communities in Social Networking Services

AUTHORS:

RICARD L. FOGUES[⋆], JOSE M. SUCH[†], AGUSTIN ESPINOSA[⋆] AND ANA GARCIA-FORNES[⋆]

{rilopez,aespinosa,agarcia}@dsic.upv.es, jose.such@kcl.ac.uk

[⋆]DEPARTAMENTO DE SISTEMAS INFORMÁTICOS Y COMPUTACIÓN

UNIVERSIDAD POLITÉCNICA DE VALENCIA, SPAIN

[†]KING'S COLLEGE DEPARTMENT OF INFORMATICS LONDON, UK

Abstract

The use of social networking services (SNSs) such as Facebook has explosively grown in the last few years. Users see these SNSs as useful tools to find friends and interact with them. Moreover, SNSs allow their users to share photos, videos, and express their thoughts and feelings. However, users are usually concerned about their privacy when using SNSs. This is because the public image of a subject can be affected by photos or comments posted on a social network. In this way, recent studies demonstrate that users are demanding better mechanisms to protect their privacy. An appropriate approximation to solve this could be a privacy assistant software agent that automatically suggests a privacy policy for any item to be shared on a SNS. The first step for developing such an agent is to be able to elicit meaningful information that can lead to accurate privacy policy predictions. In particular, the information needed is user communities and the strength of users' relationships, which, as suggested by recent empirical evidence, are the most important factors that drive disclosure in SNSs. Given the number of friends that users can have and the number of communities they may be involved on, it is infeasible that users are able to provide this information without the whole eliciting process becoming confusing and time consuming. In this work, we present a tool called Best Friend Forever (BFF) that automatically classifies the friends of a user in communities and assigns a value to the strength of the relationship ties to each one. We also present an experimental evaluation involving 38 subjects that showed that BFF can significantly alleviate the burden of eliciting communities and relationship strength.

3.1 Introduction

Social networking services (SNSs) are currently the services that are most more demanded by users worldwide. Facebook (with more than 800 million active users[1])

[1] Facebook statistics `http://newsroom.fb.com/`

and Flickr (with 51 million registered members[2]) are two of the most successful SNSs. People register to these SNSs and share images, videos, and thoughts because they perceive a great payoff in terms of friendship, jobs, and other opportunities [38]. However, the huge number of items uploaded to these SNSs and the persistence of these items in the social networks have the potential to threaten the privacy of their users[58]. For example, employers are becoming accustomed to checking the profile of the candidates in popular SNSs. If the privacy of the profile of a candidate is not properly set, what an employer sees in that candidate's profile may affect the employer's decision. It might even be possible for a stalker to infer the address of a person by looking at that person's photos posted in a social network.

Factors like the increase in public attention to privacy matters and the users' increment of familiarity with SNS technology have increased the privacy concerns in SNSs [13]. To cope with privacy threats users tend to adjust and modify the default privacy preferences set up by the SNSs since they feel that these default settings are not enough to protect them. Nonetheless, the current privacy setting mechanisms offered by SNSs seem difficult or confusing for users [94, 139]. These complications and obstacles lead to privacy policies that do not fit users' preferences, and, in turn, discourage users to show high engagement in terms of how much they participate in the SNS (e.g., the amount of photos they upload) [136].

To address this, privacy management mechanisms that are able to automate the process of privacy policy definition as much as possible are needed [39]. In this way, our long-term aim is to develop intelligent software agents that could act as privacy assistants recommending adequate privacy policies. To this aim, we need to consider the existing empirical evidence on what drives disclosure in SNSs. In particular, it has been proven that the most important factors that users consider to decide whether disclosing information is adequate or not are the strength of the relationships they have to others and the communities in which they are involved [156]. Thus, and intelligent privacy assistant agent should base its predictions on

[2] Yahoo advertising solutions http://advertising.yahoo.com/article/flickr.html

this information. The problem is that eliciting this information from the user could become a time consuming process, e.g., it would require that users specify for each of his/her friends how strong their relationship is (average user in Facebook has 130 friends [117]).

As introduced in the paper of Granovetter[56], the concept of *tie strength* defines the relationship between two individuals. In his work, Granovetter describes two different types of ties: *strong* and *weak*. On the one hand, strong ties usually include relationships such as family and close friends. On the other hand, weak ties may refer to coworkers or less trusted friends. More recent works have proposed models to predict tie strength in SNSs [53, 80, 161]. These works showed that it is possible to infer tie strength from the available personal data in a SNS. However, these works were not aimed at creating an actual tool. Thus, they only considered the predictive capabilities of the variables collected from the SNS without taking into account other factors that are of crucial importance to create an usable tool that predicts tie strength. In particular, these works did not consider either: (i) the computational cost of collecting the variables from the SNS (e.g., if the tool takes too much to complete the process this could be also seen as a time-consuming and not feasible in practice); or (ii) if all the variables apply to any possible Facebook user (e.g., these works consider language-dependent variables that could limit the applicability of their approach to speakers of other languages).

Regarding communities, they are usually defined as natural divisions of network nodes into densely connected subgroups [54]. In our context, the nodes are the contacts or friends of a given participant, and the connections between the nodes are friend relationships. Although many SNSs support some notion of community by means of allowing groups of users, they usually require that the user manually assigns each friend to the corresponding group. For example, Facebook offers the possibility of creating groups of friends, and then assigning privacy policies for each of them. However, this again has the problem that users are required to spend a considerable amount of time creating the groups and assigning friends for each of the groups (if we consider that the average number of friends in Facebook is 130, classifying all of them into groups can represent a serious challenge). Thus, the process of friend grouping should be also automatised as much as possible. To this aim, we can use one of the many existing community finding algorithms [48]. The problem is that, to the best of our knowledge, there is no empirical evidence of how this community finding algorithms perform when they are applied to real social graphs extracted from a SNS, so choosing the most appropriate one is a challenging problem.

Our main contributions in this article are:

1. We present a new tool called Best Friend Forever (BFF) that is able to automatically obtain relationship strength values and user communities from a SNS. Moreover, it allows users to further refine the results if they are not accurate enough. BFF has been implemented as a Facebook application and is publicly available at `gti-ia.dsic.upv.es/bff`.

2. We propose a new method to calculate tie strength in SNSs that considers not only the predictive capability of the variables used but also other crucial factors to develop an actual tool: the temporal cost of collecting the variables and that these variables are general enough to be applied for any SNS user. Moreover, this new method has been implemented in the tie strength module of BFF.

3. We evaluate several community finding algorithms using real social graphs (from 38 real Facebook users), and select the most appropriate one to be included in the community finding module of BFF based on their accuracy and their temporal cost.

4. We empirically demonstrate that by using BFF we are able to elicit users' relationship strength and communities requiring little intervention from users. In particular, 81.71% of tie strength values were exactly inferred and 67.08% of users' friends were correctly organized into communities.

The rest of the paper is organized as follows. Section 3.2 presents an overview of BFF and its different elements. Section 3.3 explains the tie strength prediction module and how it works. Section 3.4 presents the community prediction module and the community finding algorithms used by this module. Section 3.5 reports the results of the experimental evaluation and discusses its generalizability and limitations. Section 3.6 discusses some related works. Finally, Section 3.7 concludes the paper and outlines future research directions.

3.2 Best Friend Forever

This section introduces our tool and gives a complete overview of it. BFF aims to retrieve information from the social network that can be useful to recommend privacy policies. Specifically, the data needed is tie strength and friend groups. BFF is written in PHP and Javascript and is publicly accessible. Due to our experimentation needs, BFF is currently working as a web page; however, in the future, we plan to distribute BFF as a software program that users can execute in their own computers or on a trusted web server in order to preserve their privacy.

BFF is composed of two modules: (i) community prediction, and (ii) tie strength prediction. The community prediction module is in charge of creating chunks of users from the participant's contacts. The tie strength prediction module establishes a value of tie strength to each one of the participant's friends. In a nutshell, the input of BFF is the profile of the participant in the social network, and the output is a set of user groups and a value of tie strength for each one of the participant's contacts.

Figure 3.1 shows an overview of BFF and how it works. The interface between BFF and the user is a web page. As the figure shows, BFF collects information from the user's Facebook account. We chose Facebook as the first SNS for experimentation and for our first development of BFF due to the success of this SNS and its popularity. Nevertheless, BFF can be easily adapted to other social networks and even to social networks with a distributed architecture, like for example Friendica [3]. Therefore, before users can use BFF, they have to log in Facebook and give permission to BFF

[3] http://friendica.com/

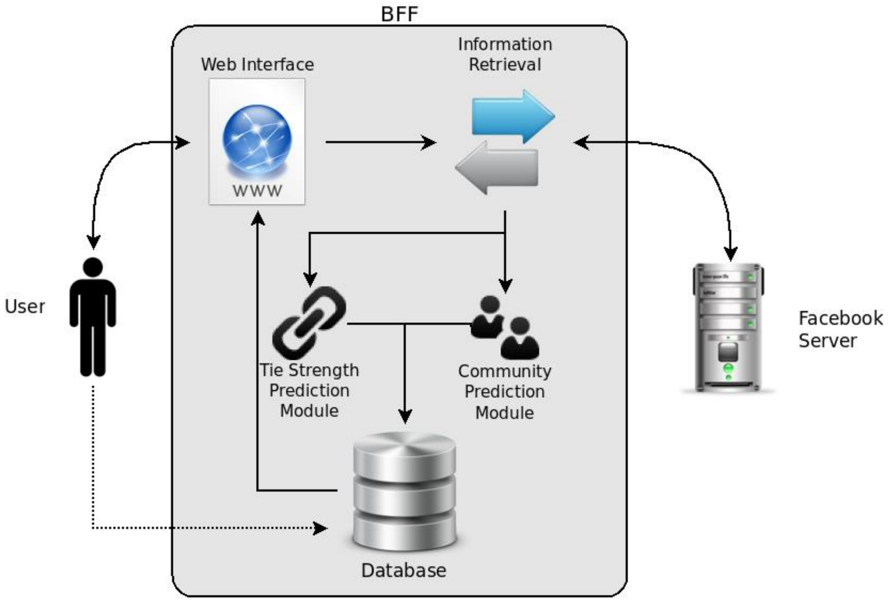
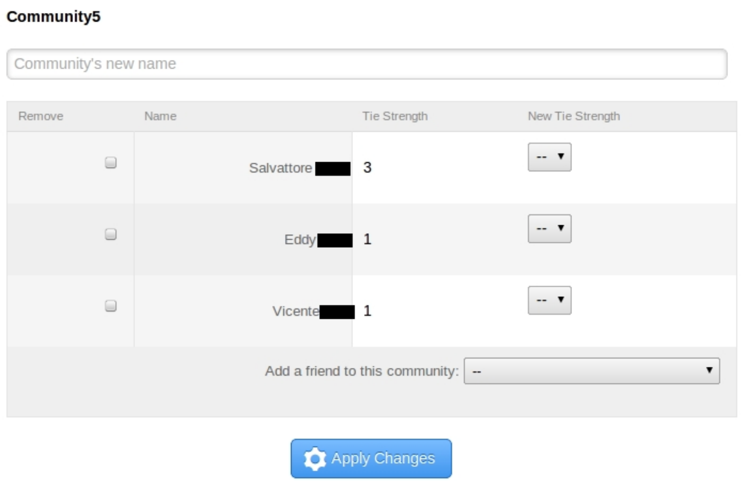


Figure 3.1: BFF Overview

to access their Facebook information. Once the permission is given, BFF requests information from the Facebook server. When all the necessary information has been collected, the information is passed to the community prediction module and to the tie strength prediction module. These modules predict a set of groups and tie strength values for the friends of the user. The predictions are shown to the user as a suggestion using again the web interface. The dotted line represents the possibility for the user to refine the suggestions created by the two modules. These modifications are stored in the database for future reference.

Figure 3.2 shows the screen where the results of BFF are presented to the user. In this example, the figure only depicts one of the communities automatically created by BFF. Part of the name of the members of the community has been hidden to preserve their privacy. As shown in Figure 3.2, the members of the community are sorted by their tie strength value. Users can refine the results by changing the name of the community, removing/adding members from/to the community, or changing the tie



Community5

Community's new name

Remove	Name	Tie Strength	New Tie Strength
□	Salvattore [redacted]	3	-- ▾
□	Eddy [redacted]	1	-- ▾
□	Vicente [redacted]	1	-- ▾

Add a friend to this community: -- ▾

Figure 3.2: Result Sample

strength value for any member in that community.

The following sections 3.3 and 3.4 explain with more details how the tie strength prediction module and the community prediction module work.

3.3 Tie Strength Prediction Module

As stated in the introduction, BFF predicts the tie strength of the relationships of the user with each person that is socially connected to her. We model tie strength as a linear combination of predictive variables. This variables are information collected from the profile user. During the creation of BFF the usability of our tool was a key factor. BFF has to be capable of predicting the tie strength accurately in a reasonable amount of time, and every user should be able to get an accurate prediction.

The selected predictive variables for BFF are based on the variables proposed in [53]. In their work, Gilbert and Karahalios propose a set of 74 predictive variables. The authors did not consider the cost of collecting the variables and their generalization, they only considered the predictive capabilities of the variables.

Before the experiment with the participants, we tested the average temporal cost of collecting the entire set of variables[4] proposed in [53]. For this test we used the profiles of 10 members of our research department. The average time of collection (without considering the cost of processing the collected information) was 1,210.73 seconds ($\sigma = 435.55$); with a maximum of more than 30 minutes for a very active profile (more than 800 friends and daily updates). We detected that the most relevant factor for the temporal cost was the number of friends. Therefore, in order to reduce the time needed for data collection, we removed variables that are collected from the profiles of the user's friends. For example, we did not collect the publications made by the participant on the walls of her friends. Instead, we only collected the publications of the participant's friends on her wall. We also removed profile dependent variables, such as home city or current job. This type of variables are usually left in blank by SNS users. This can lead to incomplete information and prediction errors [39]. We also took the decision of considering only variables that could be easily processed. More specifically, we limited the variables to those that can be simply counted. This decision provides us two advantages: the variables can be easily generalized and we largely reduce the cost of processing the collected information. For example, taking into account variables that depend on the content of a message require a natural language analyze process. Moreover, this type of variables can limit the different users that are be able to use BFF (e.g., only English speakers). Instead, counting the number of messages takes less effort and can be obtained from any user profile.

BFF collects the information from Facebook using Facebook Query Language (FQL). This language enables developers to use a SQL-style interface to query the data stored in Facebook databases. FQL queries are sent through HTTP requests. The results of FQL are paginated, thus, retrieving the entire set of elements of a specific table (e.g., photos, messages, etc.) can require several queries. Active Facebook

[4] Gilbert and Karahalios state in their paper that they consider 74 variables; however, in the paper they only show and explain 32 of these variables. In the end, we tested the information collection time considering these 32 variables.

users tend to have hundreds of friends, pictures, and hundred of posts on their walls. Therefore, collecting all the available information of the user on Facebook can take several minutes (each HTTP query takes a few seconds). We did not know the amount of information that the participants of the experiment would have on their profiles. Therefore, to avoid an excessive data collection cost, we limited the number of queries to a maximum of 20 pages per item.

Applying the restrictions explained above, we had 12 variables left. These 12 variables were not enough to cover the seven tie strength dimensions defined by Granovetter and related literature [19, 154, 93]. Specifically, two dimensions were not covered: *social distance* and *emotional support*. To cover social distance we included the variable *educational difference*. We included this variable since in the study of Gilbert [53] this variable appeared the most predictive variable in its dimension. However, as explained in section 3.5.3.1, in the end, this variable can be removed from our model since it has a very low prediction value. To cover emotional support we chose the variable *"likes"* which is not considered in [53]. Since we only counted the likes given from participants' contacts to the items collected for the other variables, this variable did not have an extra collection cost. Table 3.1 shows and explains the 14 selected predictive variables. Table 3.2 shows the tie strength dimensions and the predictive variables that belong to each dimension.

We tested the collection time for these 14 variables using the same 10 profiles of members of our research group. The average collection time was 210,48 seconds ($\sigma = 65.94$). On average, collecting this set of 14 variables was 5.77 ($\sigma = 1.28$) times faster than collecting the 32 variables proposed by Gilbert and Karahalios.

The equation below represents the tie strength s_i of the i^{th} friend. R_i stands for the vector of fourteen predictive variables of the i^{th} friend. μ_M is the mean strength of mutual friends between the user and the i^{th} friend. Finally, β is the vector of weights applied to the predictive variables and γ is the weight applied to the mean strength of mutual friends. In order to set the weight of each variable we used the findings of [53] as we wanted to avoid the use of a model that completely lacked information on

Variable	**Explanation**
Last communication	Measures the recency of the communication.
First communication	Is an approximation of the duration of the friendship.
Wall messages	Counts the number of messages exchanged using the wall.
Photos together	Counts the photos where both persons (participant and friend) are tagged.
Links shared	Counts the number web page links traded between the friend and the participant.
Initiated wall posts	Counts the number of publications posted by the friend on the participant's wall.
"Likes"	Counts the number of likes given by the friend to the participant's publications.
Inbox messages exchanged	Counts the number of private messages traded between both persons.
Inbox thread depth	Measures the length of the conversations between both persons.
Number of friends	Is the total number of friends of the friend.
Number of common friends	Counts the number of friends that are common for both persons.
Photo comments	Is the number of comments made by the friend to the photos of the principal user.
Educational difference	Measures the difference in a numeric scale: none = 0, high school = 1, university = 2, PhD = 3.
Mean tie strength of mutual friends	Taking into account the mean tie strength of the friends that are common for both persons we can capture the idea of how relationships are modified by the social cliques.

Table 3.1: Predictive variables considered.

Dimension	**Variables**
Intimacy	First communication. Number of friends. Photos together.
Intensity	Wall messages. Initiated wall post. Inbox messages exchanged. Inbox thread depth. Photo comments.
Duration	First communication.
Social distance	Educational difference.
Reciprocal services	Links shared
Emotional support	Likes
Structural dimension	Mean strength of mutual friends. Number of mutual friends.

Table 3.2: Predictive variables and tie strength dimensions

the relative importance of each variable to predict tie strength.

$$s_i = \beta R_i + \gamma \mu_M$$

$$M = \{s_j : j \text{ and } i \text{ are mutual friends}\}$$

After collecting the predictive variables for the friends of the user, the variables are normalized. Then, the tie strength is calculated for each user. The results are normalized to a numeric scale 1-5, where 1 represents that both persons are very distant (mere acquaintance) and 5 that they are very close. The results are presented graphically, as shown in Figure 3.2, so that users are sorted by group and by tie strength. It is easier to figure out the value of the tie strength of a person by comparing it to the values of the tie strength of the relationships with others. As in the grouping step (explained below), the participant can refine the results of the tie strength calculation.

3.4 Community Prediction Module

The community prediction module is in charge of dividing the network of relationships of the user into communities. This module queries Facebook about the friends of the user and the friends of those friends (mutual friends). With this information the module builds a graph where the nodes are the friends of the user, and the connections between the nodes are friend relationships. This graph is used as the input for the community finding algorithm. The output of the algorithm, a partition of the graph, is shown to the user. The user can modify the communities proposed by the algorithm.

As in the tie strength prediction algorithm, we wanted to present to the participants of the experimental evaluation a suggestion that they can modify. Creating every community from scratch can be a challenging task and we wanted to avoid participants of getting tired of the experiment. The algorithm proposed by Shen et al. in [126] was chosen as the initial community finding algorithm for BFF. The algorithm is founded on the triadic closure principle, which suggests that, in a social network, there is an increased likelihood that two people will become friends if they have friends in common. Based on the results obtained by the authors, this algorithm performs accurately on natural created communities which is the type of communities that the community prediction module has to manage.

According to the results of Shen et al. [126], their algorithm performs better than Infomap [123] and Louvain [10] algorithms. These two community finding algorithms (Infomap and Louvain) are two of the best ones [88]. On the one hand, Infomap uses the probability flow of random walks on a network as a proxy for information flows in the real system and decompose the network into modules by compressing a description of the probability flow. On the other hand, Louvain algorithm is founded on a heuristic method that is based on modularity optimization. The modularity of a partition is a scalar value between -1 and 1 that measures the density of links inside communities as compared to links between communities [54].

As we did not know if the Shen's algorithm was going to be accurate with Facebook communities, we also tested Infomap and Louvain algorithms. The results of the test comparing the three algorithms are shown in the evaluation subsection 3.5.3.2.

3.5 Experimental Evaluation

The goal of our experimental study is to evaluate the accuracy of our BFF tool in terms of community and tie strength prediction. Specifically, we want to answer the following questions:

- How accurate is the community module in grouping the contacts of a user?
- How accurate are the predictions of the tie strength module?
- Do users perceive that BFF is a good tool in general? In other words, do they think that BFF is capable of inferring accurate information from their available data on Facebook?

To answer these questions, we performed an experimental evaluation with Facebook users. In the rest of this section, we firstly introduce the experimental settings and, after that, we report the main findings.

3.5.1 Participants

Our 38 participants were recruited using a Facebook page as well as posters posted on the Universitat Politècnica de València university. We used the viral properties of publications on Facebook to attract participants out of the college environment. Participants were rewarded with a gift voucher for El Corte Ingles (a famous chain of shopping centers in Spain).The participants also entered into one Nexus7 tablet raffle.

The participants filled a form with demographic information. The sample consisted of 9 women (24%) and 29 men (76%). 52% of the participants had an age between 18 and 24, 25% had an age between 25 and 29, 22% had an age between 30 and 39, and 1% of the participants had an age between 40 and 49. Regarding studies degree, the majority of them had a college degree (71%), 14% of them had a PhD, other 14% of participants had high school degree and 1% of the participants had a primary school degree. Finally, 76% of the participants were students and the other 24% were working. We consider that the Facebook viral effect succeeded as at least every age, study degree and professional status was represented. This is specially important in the academic environment, when very often, experimentation with humans tend to be limited to college students. However, attending to the study made by Johnson et al. [76], two demographic groups were underrepresented, the group of women and the age group 40+.

Regarding the number of participants' friends on Facebook, the minimum number of friends was 35 and the maximum was 529. The mean number of friends per participant was 232.19. In total, we analyzed 12803 friend relationships. The majority of participants use Facebook regularly, 84% of them enter Facebook several times per day, the other 16% visit Facebook at least once every few days.

3.5.2 Method

The participants in our experiment had to try BFF and evaluate its precision. BFF was created to ensure that its use would be easy for anyone. The participants only had to access to the web page of BFF, log in with their Facebook account, and start the application.

After BFF completed its process, the participants were requested to correct any possible errors in tie strength prediction and in user grouping. Users could change the tie strength value of any contact, move users freely from one community to another, and create new communities. These possible corrections were stored in order to evaluate the performance of BFF.

Finally, the participants were requested to answer a short survey to find out their opinion about BFF. The survey was composed of the four following questions:

1. How well did BFF group your friends into communities?

2. How well did BFF predict the tie strength between you and your friends?

3. In general, how accurate do you think BFF is?

4. How accurate do you think BFF is considering it only accesses your information on Facebook? For example, if one of your friends on Facebook is your brother, but you have never interacted with him on Facebook, it is impossible for BFF to accurately predict the tie strength between you and your brother.

Each question was rated on a Likert scale 1-5: 1 = very bad, and 5 = very good. The first and second question addressed specific parts of BFF (the grouping feature and the tie strength prediction respectively). The third and fourth questions were general questions. The intention of the fourth question was to clarify the limitations of BFF to the users. Currently, BFF is limited to the bounds of Facebook; therefore, it only considers the interactions and social connections that occur on Facebook. In future work, we expect to collect information from different sources than Facebook, so BFF will be able to avoid this limitation.

3.5.3 Results

In this section we analyze the results obtained for both modules, tie strength prediction and community prediction. Apart from reviewing the performance of both modules, we also study how their performance can be enhanced.

Variable	β
Last communication	-1.3487
First communication	1.2603
Mean strength of mutual friends	1.0546
Photos together	0.9529
Likes	0.6223
Number of friends	-0.4903
Inbox thread depth	0.4785
Initiated wall posts	0.3138
Inbox messages exchanged	0.2602
Links shared	0.1282
Wall messages	0.1219
Number of common friends	-0.0338
Educational difference	-0.0164
Photo comments	-0.0028

Table 3.3: β coefficients for predictive variables after regression

3.5.3.1 Tie Strength Prediction Module

With regard to tie strength prediction, the module performed very accurately. It achieved a Root Mean Squared Error (RMSE) of 0.6271 and a Mean Absolute Error of 0.1791 on a discrete scale 1-5, where 1 is the weakest and 5 is the strongest. We chose to discretize[5] the tie strength in order to facilitate the understanding of the results to the users.

We performed a linear regression analysis to observe what variables were more useful for tie strength prediction and to inspect if the initially chosen coefficients were suitable. The β coefficients for the variables are shown in Table 3.3.

[5] The discretization process might have caused a higher prediction error. For example, a user with a tie strength of 3.6 and another with a strength of 4.4 will be both assigned a strength of 4 during the discretization process. As future work, we plan to study the effect of discretization in the prediction error, so that we could achieve a trade-off between the understandability of the results and the error introduced because of the discretization.

Using a model with the coefficients specified in the Table 3.3 we obtained a RMSE of 0.614. The small difference between the error of the previous model and the model after the regression process shows that the initial coefficients applied to the variables were appropriate.

Table 3.3 also shows some variables with very low coefficients. It is interesting to study if it is possible to remove these variables from the model without degrading the predictions significantly. As explained before, each variable requires querying the SNS for the information. This process can be time consuming, specially on very active users. Reducing the amount of information needed for tie strength prediction directly improves the time performance of the module. We performed a multilinear stepwise[6] regression to observe what variables could be removed from the model maintaining an acceptable error rate. Table 3.4 shows the predictive variables sorted by β coefficient after the stepwise regression. As it can seen on the table, the variables with lower β coefficient are also the variables with higher p-value. Therefore, we can create a new model without these variables (*links shared, wall messages, photo comments, common friends*, and *educational difference*) and still maintain an acceptable level of accuracy. This new model, with only 9 variables, has an RMSE = 0.7964. Comparing this error rate to the error rate obtained by the complete model, the increase is not significant. Besides the performance benefit, as explained above, removing education degree is specially beneficial as profile dependent variables, like this one, are usually left in blank by SNS users. This can lead to incomplete information and prediction errors [39].

As stated previously, SNS users struggle to set up privacy settings. If the aim of BFF is to lighten the burden of this task, its suggestions cannot contain a huge number of errors that need correction. Therefore, it is crucial to keep to the minimum the number of corrections that the users need to make to the suggestions presented by BFF. Analyzing the experimental data, we found that the participants made an average of 42.47 tie strength corrections. Considering that the average number of friends of

[6] A sequence of F-tests is used to control the inclusion or exclusion of variables

Variable	β	**p-value**
Last communication	-1.3527	< 0.001
First communication	1.2684	< 0.001
Mean strength of mutual friends	1.0485	< 0.001
Photos together	0.9608	< 0.001
Likes	0.6817	< 0.001
Inbox thread depth	0.5110	< 0.001
Number of friends	-0.4986	< 0.001
Initiated wall posts	0.3412	< 0.001
Inbox messages exchanged	0.2675	< 0.001
Links shared	0.1836	0.1036
Wall messages	0.1791	0.1023
Photo comments	0.0562	0.62
Number of common friends	-0.0303	0.55
Educational difference	-0.0161	0.5161

Table 3.4: β coefficients and p-values for predictive variables after stepwise regression

our participants was 232.19, having to correct only 18.29% of friend relationships can effectively speed up the process of classifying friends before setting privacy policies. It is worth noting that BFF only needs 14 predictive variables, which ensures that it can work fast while maintaining a good accuracy. Moreover, the generalization of the predictive variables allows BFF to be accurate without taking into account the specific characteristics of the user.

3.5.3.2 Community Prediction Module

In order to evaluate the performance of the community prediction module we used the Normalized Mutual Information (NMI). NMI gives a measure of the quality of the partition obtained by the module comparing that partition to the partition made by the user. NMI is in range [0,1], it equals 1 when both partitions are equal and 0 when both partitions are independent. We use the method proposed by Lancichinetti et al [89] to calculate the NMI. This method defines the NMI between two partitions

X and Y as:

$$N(X|Y) = 1 - \frac{1}{2}[H(X|Y)_N + H(Y|X)_N]$$

Where $H(X|Y)_N$ and $H(Y|X)_N$ are the normalized conditional entropy. To calculate this entropy, the algorithm considers the differences between the most similar clusters in both partitions. In other words, the algorithm calculates the total entropy between each cluster in partition X and the most similar cluster in partition Y, and vice versa.

As explained in section 3.4 the participants were asked to apply any needed correction to the communities created by Shen's algorithm. Besides this algorithm, we chose two additional algorithms (Infomap and Louvain) to compare the performance of all three. We calculated the NMI between the communities created by each algorithm and the final communities specified by the participants. Figure 3.3 shows the mean NMI obtained by each algorithm. Infomap is the algorithm with best performance as it achieves a mean NMI of 0.55. Infomap yields more and smaller communities than the other two, this behavior seems to coincide more precisely with the natural partitions made by the participants. Since the users were asked to correct the results obtained by Shen's algorithm, rather than creating their communities from scratch, some bias might have been introduced. It is likely that some communities created by Shen's algorithm that were slightly different to the preferences of the participants were considered as correct by the participants. It is possible that if Infomap had been chosen as the initial community finding algorithm, its NMI would have been even better. In this case, the participants would have corrected the suggestions offered by Infomap. Hence, a similar bias as the one introduced in the experiment towards Shen's algorithm would have been introduced towards Infomap algorithm.

In order to maintain BFF usable it is important to keep the number of corrections that users need to make to BFF suggestions to the minimum. We analyzed the number of changes that participants needed to make to the communities suggested. Figure 3.4 shows the mean proportional number of changes made to community suggestions. As can be inferred from the average NMI obtained by the three algorithms, Infomap

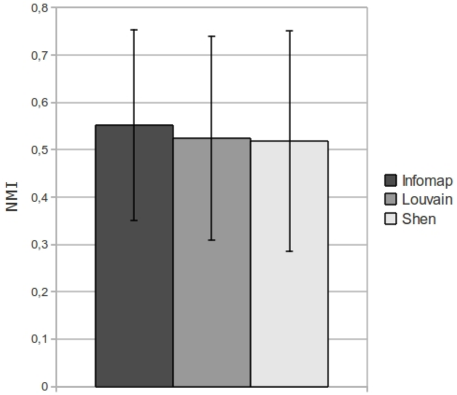


Figure 3.3: NMI for different community finding algorithms

outputs needed less corrections than Shen's and Louvain's. In average, participants needed to move 32,92% of their contacts to other communities in order to adapt the partition calculated by Infomap to their preferences.

Another factor to consider when selecting the community algorithm is the execution time. We performed a test to measure the execution time for each algorithm using the evaluation data. Figure 3.5 shows the average execution time for each algorithm in milliseconds. The fastest algorithm is Louvain; however, every algorithm is very fast and their execution time do not affect the general performance of the tool.

As a conclusion, Infomap algorithm is the best algorithm overall. It achieved the highest mean NMI and its outputs required less corrections than Shen and Louvain algorithms. Infomap is not the fastest one, however, the execution time is not relevant as the three algorithms are very fast with the size of Facebook communities and the execution time differences are not significant. Therefore, we have replaced the Shen algorithm with Infomap algorithm for future versions of BFF.

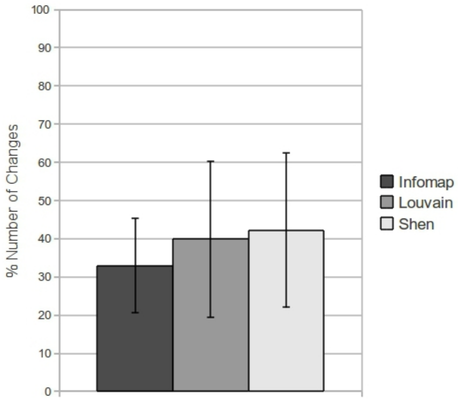


Figure 3.4: Mean proportional number of community changes

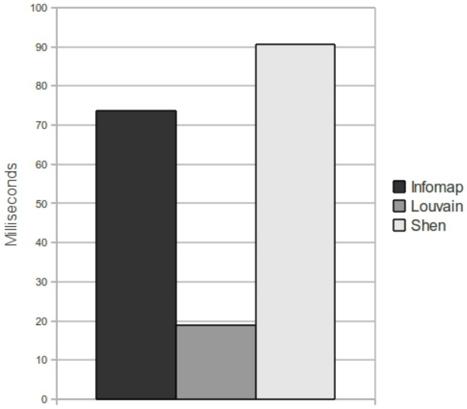


Figure 3.5: Average execution time for each algorithm in milliseconds

3.5.3.3 Survey results

The participants also rated the performance of BFF by answering a short survey. Each question was rated using a Likert scale 1-5. Figure 3.6 shows the mean rating for each survey question. The results reflect that the participants rated BFF performance positively. The participants perceived a slightly better accuracy in tie strength prediction than in community prediction. This perception can improve after the replacement of the community finding algorithm , as the the Infomap algorithm has proved to perform better. Another result to note is that the participants rated the second general question (question 4) higher than the first general one (question 3). When answering the first general question, the participants did not consider the limitations of BFF. Therefore, even when almost every friend was rated correctly, they detected mistakes. Due to the brief explanation in the second general question about how BFF works, the participants realized that BFF is limited by the bounds of Facebook, and, for example, that it cannot predict the tie strength of a relationship that mainly occurs outside Facebook. When the participants became aware of the limitations of BFF, they took into account how they interacted with othersThey achieved an accuracy of 86% on Facebook in order make their judgments. This explains the better rating for the second general question.

3.5.4 Generalizability and Limitations

Since Facebook is no longer limited to this group of users, it is necessary to take into account the participants of other demographic groups. Despite the fact that the group of women and people 40 years old or older were underrepresented, we believe that our sample accurately reflects the current demographics of Facebook [76].

Another element that increases the generalizability of our study is that it is not based on surveys. The utility of surveys is unquestionable; however, the results obtained through surveys can introduce bias. The data analyzed in this paper is real data retrieved from real Facebook users. We collected a large volume of information

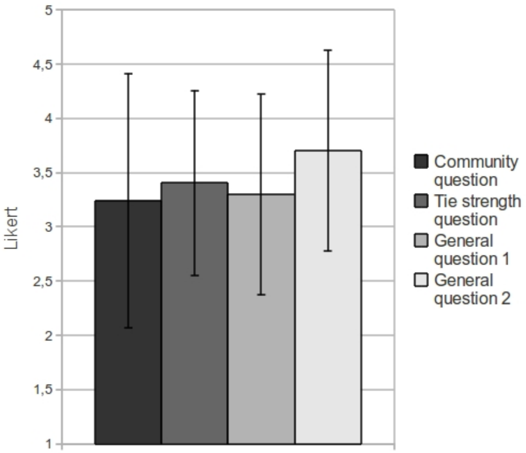


Figure 3.6: Score for the survey questions

that is not possible to obtain through surveys (thousands of relationships and their characteristics).

It is possible that our method introduced bias. Participants of the experiment were asked to correct the results inferred by BFF, thus, their perception of their relationships could be affected by the results shown. The tie strength values specified by the users could be modified by the results shown. Also, as explained above, the users were asked to correct the results obtained by Shen's algorithm, rather than creating their communities from scratch. Despite the possible bias introduced, the aim of BFF is providing good enough recommendations that help users to classify their relationships with others on a SNS. Therefore, the measure of the performance of BFF has to be based on how appropriate the recommendations of BFF are for users and the number of corrections that the users need to perform to these recommendations.

3.6 Related Work

Recent works have proposed models to predict tie strength. Gilbert et al.[53] proposed a model, based on Granovetter's work, that predicted tie strength among the users of Facebook. The authors identified a set of 74 predictive variables that can be found on Facebook. They performed an experiment to infer the tie strength of a portion of the participants' friends. They achieved an accuracy of 86%. Another work that predicts tie strength of social links is [80]. Like in the work of Gilbert, the authors define a set of 50 predictive variables. In this work the authors aim to discriminate strong links from weak links. However, they do not consider a scale in the strength of the link, they are either strong or weak. These two works use a supervised learning model that needs human intervention to work properly. Aiming at the same objective, Xiang et al.[161] proposed a model to infer relationship strength based on profile similarity and interaction activity, with the goal of automatically distinguishing strong relationships from weak ones. It is worth noting that this model relies on an unsupervised learning method, but it lacks a empirical evaluation with real users. All three works show that it is possible to infer tie strength from the available personal data in a SNS. These three works differ from ours in that they aim to create models to predict tie strength from the information available on a SNS. However, they do not offer tools that social network users can use to help them to form friend groups and set privacy policies. Moreover, they only consider the predictive capabilities of the variables chosen for their models, but they do not take into account factors like the computational cost of collecting these variables, which is an important factor when creating an usable tool.

The other main feature of BFF is that it suggests friend groups to the participant user. The main idea is that with the grouping and tie strength information the user has enough elements to create appropriate privacy policies. The work of Fang and coworkers [39] proposes a tool that suggests privacy policies for certain elements of a Facebook user profile. This work bases the privacy suggestions in grouping user's contacts in contexts. Every contact in the same context is granted the same access permissions. The authors present a tool called Privacy Wizard that helps user to set the privacy policies to protect user's traits, like birth date, address, and telephone number. This work does not consider tie strength, and as the authors proved in [156], it is a key variable to consider when determining the disclosure degree of the elements being shared in a social network.

Other works present mechanisms that can partially infer users' social network and its characteristics from sources of information different than SNSs. In [30] the authors propose a method that extracts a social network for a user given her mailbox and the information available on Internet. A similar approach is presented in [102]. In this work the authors present Polyphonet. From a given set of persons, the authors find the social connections among them by querying to Google. The authors estimate the strength of the relationship between two persons by co-occurrences of their two names. These two works differ from ours in that they do not rely in a SNS to extract social information from users. However, this approach also has limitations. Relying on information sources that do not necessarily contain social relevant information may lead to errors. For example, two persons may appear in several web pages together but do not have any social link. In order to avoid this problem, both works ([102, 30]) require a predefined set of persons that will form the social community. In contrast, relying only on Facebook data guarantees that the social links will actually exist, but may also lead to errors. Even when the connection truly exists, the interactions between two persons may occur outside Facebook. Therefore, the strength of such link will be incorrectly predicted by our software. In the future, we plan to expand the search of variables for defining the groups and the tie strength with information that can be found outside the social network, like the information available in the participant's mailbox or in the personal web page of a user of the social network.

The work of Murukannaiah and Singh [105] presents Platys Social. The authors developed a software that runs on a mobile device. This software learns a user's social circles and the priority of the user's social connections from daily interactions. The

Feature	Accuracy
Tie strength prediction with 14 variables	81.71%
Community prediction	67.08%

Table 3.5: Results overview.

software infers the interactions from information that is available on mobile devices, such as wi-fi networks, bluetooth connections, phone calls, and text messages. The work of Murukannaiah and Singh presents a new approach for extracting social information from the real world, and not only from Internet. Their work and ours could be merged so that tie strength could be computed taking into account day by day encounter frequency and the information stored on a SNS like Facebook.

3.7 Conclusions and Future Work

In this paper, we have presented the BFF tool that is able to automate the elicitation of tie strength and user communities to a large extent. In particular, we evaluated it using real-world data from real users of Facebook. BFF achieved 81.71% accuracy in tie strength (i.e. users only needed to correct 18.29% of their relationships). Moreover, the tie strength prediction model used by BFF was composed of only 14 variables, and it could be even simplified to use 9 variables with a slight loss of precision.

Regarding community prediction, BFF achieved 67.08% accuracy (i.e. users only needed to move 32.92% of their contacts to different communities). We evaluated three different community finding algorithms (Infomap, Louvain and Shen) in order to find the best one for the objective of BFF. Infomap algorithm outperformed the other two algorithms achieving a better accuracy. Finally, according to a survey performed by the participants in the experiments, we obtained that users consider BFF as providing good predictions of tie strength and communities. Table 3.5 shows an overview of the results.

Many research paths open from here. The first one, and the motivation of this work, is to develop intelligent personal agents that will aid users in the definition of their privacy policies for SNSs. To this aim, this intelligent agent will use the extracted information to recommend adequate privacy policies. Another path for further research is to improve the predictive capabilities of BFF by collecting information from other sources than the SNS (such as users' mailbox, personal web pages, Internet search engines and mobile devices [30, 102, 105]).

Finally, it is worth noting that the information that our tool provides can also be used in many other agent-based applications. For instance, the tie strength among agents is needed to obtain the optimal social trust path in complex social networks [97]. Moreover, in automated negotiation environments, agents could judge the outcome of a negotiation as being distributively fair based on the tie strength between them [129]. Apart from being used in agent-based applications, BFF could also be used in many other more general applications. For instance, it can be very useful to perform experiments with humans in which either tie strength or user communities are needed to evaluate the results of the experiments (such as in [156]). In this case, BFF can speedup the experiments by automating part of the process for eliciting tie strength and user communities from the participants.

CHAPTER

4

Exploring the Viability of Tie Strength and Tags in Access Controls for Photo Sharing

AUTHORS:
RICARD L. FOGUES[⋆], JOSE M. SUCH[†], AGUSTIN ESPINOSA[⋆] AND ANA GARCIA-FORNES[⋆]
{rilopez,aespinosa,agarcia}@dsic.upv.es, jose.such@kcl.ac.uk
[⋆]DEPARTAMENTO DE SISTEMAS INFORMÁTICOS Y COMPUTACIÓN
UNIVERSIDAD POLITÉCNICA DE VALENCIA, SPAIN
[†]KING'S COLLEGE DEPARTMENT OF INFORMATICS LONDON, UK

Abstract

Social Network Sites (SNSs) such as Facebook and Google+ allow users to store and share large collections of photos. SNSs offer access controls that protect those photos from unwanted audiences. However, due to the lack of usability of these access controls, people struggle to configure them. First, we collected sharing policies for photos in a study with 34 Facebook users. Then, we define three metrics that enable researchers to evaluate the ease of use and complexity of access controls for photo sharing, and, employing the data collected in the study, we evaluate 15 access controls, each one with a different combination of attributes. The results obtained show that an access control that takes into account groups, tags, and the tie strength of relationships can be managed more easily than current approaches, reducing the burden of configuring the privacy settings for photos on SNSs.

4.1 Introduction

Photos are one of the most common items uploaded to SNSs. People share photos on SNSs because they perceive a great pay-off in terms of friendship and other social opportunities [38]. However, since photos can disclose sensitive information, they require complex privacy settings [6]. In order to control how photos are disclosed, SNSs offer basic access controls [23]. These access controls are based on groups of contacts. For example, Facebook has friend lists and Google+ offers friend circles. Users can employ these groups of contacts to specify who is allowed to access a piece of content.

Although access controls of SNSs aim at simplicity, users struggle to manage and understand them [94]. To improve the understandability and usability of access controls, research works propose adding different attributes to access controls in SNSs [23, 25]. These attributes include variables such as social distance, tie strength, or groups of contacts. However, few works evaluate what combinations of attributes can potentially cover the privacy preferences of the users with the lowest possible complexity. In this study we evaluate the effects of the addition of new attributes (and combinations of them) to photo access controls. Our contribution is twofold: we (1) propose three quantitative metrics that enable us to evaluate the impact of new attributes on current access controls.; and (2) employing the proposed metrics, we evaluate the performance of 15 combinations of new and current attributes for access controls.

4.2 Related Work

Yeung et al. [163] prototyped the management of privacy for photos that considers content type. Hart et al. [61] proposed a mechanism to manage privacy for blogs based on tags. Their mechanism enables users to define groups manually or to group potential viewers by attributes that they all share (e.g., workplace or same school). The main focus of their study is to compare basic sharing policy mechanisms for blogs with a tag-based approach. The authors did not use real data from the participants, instead, they created artificial data for imaginary users and asked the participants to manage that data as if it was theirs. Thus, they do not examine users' actual preferences, as we do. Their results show that an approach that uses tags is more usable than one that does not. However, they do not evaluate if other combinations of attributes can further improve the performance. Squicciarini et al. [134] proposed a sharing policy recommender tool which considers tags and contact groups. Since they aimed at accurately inferring new sharing policies, they did not evaluate the effect of tags as a new attribute for an access control. Further, they evaluated the performance of their tool using predefined photos, instead, we use personal photos from the participants. Klemperer et al. [83] evaluated the usability of an access control based exclusively on tags. The authors aimed at evaluating whether tags can be used to organize photos and define their privacy at the same time. However, they did not compare the performance of their access control with

SNSs' current approaches, or with other access controls with different attributes.

4.3 Attributes of Access Controls

The attributes evaluated in our study are the following:

Tags (*Tag*): The different categories that define the content of the photo.

Communities (*Com*): Groups of contacts created by the users.

Tie Strength (*Tie*): The individual tie strength that users have with each one of their contacts based on a Likert scale from 1 to 5 (1 = weak tie, 5 = strong tie).

Individual Contacts (*Ind*): The current access control of Facebook allows users to specify individual contacts in sharing policies. We also consider this attribute for the different access controls that we evaluate in this section.

We use a combination of abbreviations of the attributes to name an access control. For example, the name of an access control that takes into account tags and tie strength is *TagTie*.

4.4 Method

Our investigation is based on real data retrieved from Facebook users. The information that we needed for our study is divided into three types: (i) the characteristics of the relationships between the participants and their contacts, (ii) the sharing policies that the participants apply to the photos on their Facebook accounts, and (iii) the tags of these photos.

4.4.1 Participants

The sample consisted of 11 women and 22 men. The mean age of the participants was 25.66 ($SD = 6.19$) with 18 years as the minimum age and 45 years as the maximum age. For education level, the majority of them had a college degree (20), 6 of them had a PhD, another 6 of participants had a high school degree, and one participant had a primary school degree. Finally, 76% of the participants were students and the other 24% were working.

4.4.2 Collection of Social Data

To collect the information of the relationships of the participants, we used the BFF application [41]. This is a Facebook application that helps users organize their relationships. BFF automatizes the process of friend grouping and tie strength definition. BFF collects predictive variables from the user's profile. These variables include data such as the number of messages exchanged with a friend, who appears in the photos of the user, or the total number of friends. With the collected data, BFF infers tie strength values and friend communities. BFF represents tie strength on a Likert scale 1–5 (1 = minimum, 5 = maximum). The results yielded by BFF can be refined by the users. Communities were not exclusive, one single contact could be included in several communities if the participant considered it appropriate. In total we collected 735 communities and the average number of communities per participant was 22.27.

4.4.3 Definition of Sharing Policies and Photo Tags

The participants defined sharing policies for their photos on Faceboook. First, our application collected the photos of the participants from their Facebook profiles. The photos were sorted and organized using the same album structure that the users have on their profiles. Since the access control of Facebook is based on individual contacts and contact groups, during this step, the participants defined sharing policies using the same groups and individuals that they corrected or created during the previous part of the study. Participants were asked to define their ideal policies for individual photos, and they created as many as they found necessary. The application we built enabled participants to assign the same policy to every photo in the same album if they found that appropriate. As in Facebook, participants were told that blocking takes precedence over granting access.

A majority of photos (64.89%) were assigned a public policy. On average, each photo was accessible by 89.04% of contacts. However, if we analyze only the photos that did not have a public policy we observe that the defined sharing policies were somewhat restrictive. On average, 54.11% of contacts could access every photo with a non-public policy. We found that 4 male participants used only public policies for their photos. Since this study is focused on how users define sharing policies, we do not consider their information for the evaluation of the access controls.

Participants were required to classify each album with one or more tags after defining their ideal sharing policies. The predefined tags were: family, close friends, colleagues, party, kids, travel, animals, self-portrait, fun, artistic, and other. This set of tags was extracted from the most common[1] tags used in Flickr the popular photo-exchange social network[2]. Overall, 9% of the photos were classified with *other* as one of their tags. However, only 1.13% of the photos were classified exclusively as *other*, thus, the predefined tags covered almost the entire set of photos of all participants. We told the participants that they had to assign tags for the content of the photos, not for their appropriate audience. To minimize the risk of participants using tags to define audiences, we deliberately set the tagging task after the sharing policy definition task.

[1] Even though the tag *colleagues* is not a popular tag in Flickr, we added it because we felt it makes sense in a friendship-focused social network such as Facebook.

[2] `http://www.flickr.com/photos/tags/`

4.5 Evaluation

In total, we defined 15 different access controls using the different combinations of the attributes explained above. We aim at comparing the performance of these 15 access controls. However, asking the participants to define their privacy preferences with 15 access controls would not have been feasible in terms of time and task complexity. Therefore, to find how the policies defined by the participants when using each access control model might look like, we use decision-tree classifiers. These classifiers automatically generate a representation (a tree of rules) of how users could use the available attributes in each access control model to define sharing policies that match their privacy preferences. Since the rules generated are, somewhat, simulations of participants' general privacy preferences, they do not capture every detail. For example, a participant may share every family photo with his father except a photo that depicts the preparation of a surprise birthday party for his father. Our method is not able of capturing this nuance, and the birthday preparation photo will be incorrectly classified (his father will be allowed to see it).

For the creation and evaluation of the decision-tree classifiers we used the C4.5 decision tree [118] and its implementation in the Weka[3] data mining tool. All the classifiers considered two classes: allow or deny. For each photo and contact, the classifier had to decide if the photo could be accessed by that contact (allow) or not (deny). The number of dimensions of the feature vector of each classifier depended on the attributes considered by the access control. For example, if the access control takes into account tags, communities, and tie strength, then, the vector of features will have a dimension for each possible photo tag, each community defined by the participant, and one dimension for the tie strength.

[3] http://www.cs.waikato.ac.nz/ml/weka/

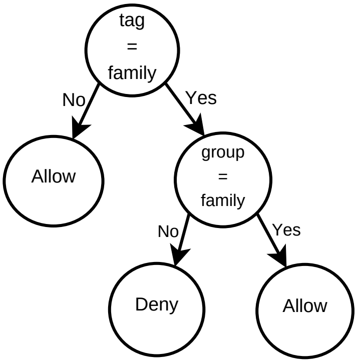


Figure 4.1: Example tree.

4.5.1 Metrics

Coverage: This measures how well the privacy preferences are correctly represented by the automatically created rules. In other words, for each pair (photo and contact) we checked whether or not the tree generated for that specific access control classified correctly the pair into one of the two classes (allow or deny). The coverage gives us information about how well the access control can express the actual privacy preferences of the user.

Number of rules: To find an appropriate access control, it is necessary to consider the trade-off between coverage and complexity. If an access control requires many rules to cover the privacy preferences of the user, it is likely that the access control will have a low usability. To measure the number of rules generated by each access control, we counted the number of leaves in the generated tree. For example, Figure 4.1 shows a simple example tree. In this tree, there are three leaves, thus, the access control has three rules: (i) family members can see family photos; (ii) non-family members cannot see family photos; and (iii) non-family photos are public.

Complexity level: Similar to the number of rules, rule complexity can be detrimental to the usability of the access control. The level of complexity of a rule is given by the number of tags, community identifiers, tie strength thresholds, and individual contact identifiers utilized. In the example shown in Figure 4.1, there are two rules

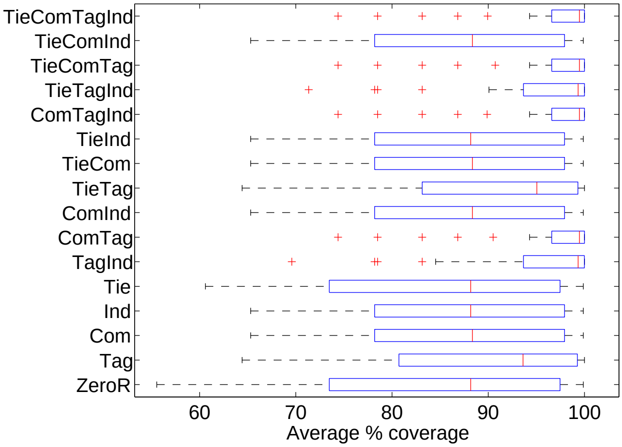


Figure 4.2: Coverage of access controls created with different attribute combinations.

with complexity 2 (the community and tag attributes are used), and one rule with a complexity of 1 (only the tag attribute is used).

4.6 Results

Figure 4.2 shows the coverage obtained by the different combinations of attributes. We use the coverage obtained by the*ZeroR* classification method as a benchmark for the different access controls. This method is the simplest classification method since it relies on the target and ignores all predictors. The *ZeroR* classifier simply predicts the majority class.

According to a Lilliefors test [92] with a 95% confidence interval, the coverage values obtained by the different access controls come from a normally distributed population. Therefore, to test whether or not the differences in coverage were significant, we performed a series of t-tests with a 95% confidence interval. Since the data used for all the access controls was the same, we used paired t-tests. To counteract the problem of multiple comparisons, we applied Holm-Bonferroni correction to the series of t-tests. The test shows that some differences are not statistically significant. The statistical differences show four different groups of access controls. Specifically, the group of access controls with the highest coverage is formed by: TieComTagInd, TieComTag, ComTagInd, ComTag, and TagInd. The second group is formed solely by TieTagInd. The third group contains TieTag and Tag. Finally, the fourth group is formed by the rest of access controls: TieComInd, TieInd, TieCom, ComInd, Tie, Ind, and Com. Overall, according to these results, tags improves the coverage of an access control the most, followed by communities, individuals, and tie strength, which affects coverage the least.

Figure 4.3 shows the number of rules generated by each access control. The results show that there were big differences in the number of rules generated. Several access controls require a number of rules that make them unmanageable for a human user. The access control with the least number of rules was *Tie*. Actually, this low number of rules explains its poor performance; the trees generated with this access control were almost the same as the rules generated by the *ZeroR* classifier. Among the access controls with good performance, *ComTag* has a slightly lower average number of rules.

Figure 4.4 depicts the complexity levels of the rules generated by the access controls. In general, the median level of complexity was around 4. Obviously, the lack of expressivity of *Tie* limited the complexity of this access control. *Ind*, *TieInd*, *TieTagInd* and *TagInd* produced also rules with a low level of complexity; however, as shown in Figure 4.3, they generate a large number of rules. Overall, *ComTag*, *ComTagInd*, *TieComTag*, and *TieComTagInd* offered the most balanced results: good coverage, a small number of rules, and low levels of complexity.

4.7 Discussion and Conclusions

We propose 15 different acces controls and three metrics to evaluate them. Using real privacy preferences, we compare the performance of these 15 access controls. According to our results, an access control that takes into account tags, communities, and tie strength requires a low number of rules with low complexity to express the

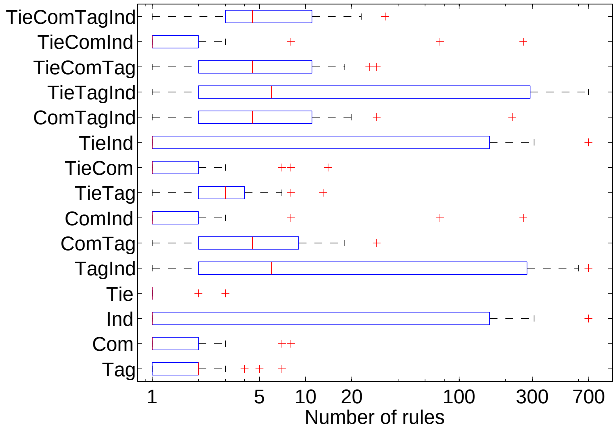


Figure 4.3: Number of rules generated by each access control.

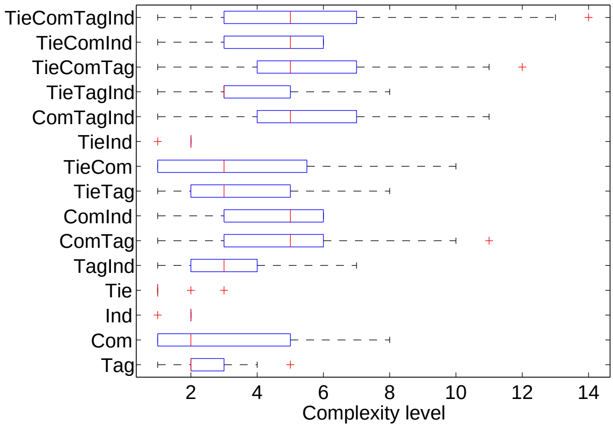


Figure 4.4: Complexity level of the rules generated by each access control.

general privacy preferences of the users with coverage that is good enough.

Analyzing the two new attributes individually, on the one hand, the results obtained in our study point out that access controls with the attribute tags achieve good coverage with low complexity. This shows that tags play a key role during sharing policy definition.

On the other hand, the tie strength attribute does not show an impact on access controls as positive as tags. One of the reasons behind this could be that participants did not assign tie strength values depending on how much they share on the SNS but outside of it. Future work should investigate whether users employing an access control that uses tie strength as a means to define sharing policies create less complex and more accurate policies than those who use an access control without it.

CHAPTER

5

Tie and Tag: A Study of Tie Strength and Tags for Photo Sharing

AUTHORS:
RICARD L. FOGUES*, JOSE M. SUCH†, AGUSTIN ESPINOSA* AND ANA GARCIA-FORNES*
{rilopez,aespinosa,agarcia}@dsic.upv.es, jose.such@kcl.ac.uk
*DEPARTAMENTO DE SISTEMAS INFORMÁTICOS Y COMPUTACIÓN
UNIVERSIDAD POLITÉCNICA DE VALENCIA, SPAIN
†KING'S COLLEGE DEPARTMENT OF INFORMATICS LONDON, UK

Abstract

Tie strength and tags have been separately suggested as possible attributes for photo access controls in Social Network Services. However, an evaluation is missing about the benefits/drawbacks of adding one or both of these attributes to the ones already used in access controls for Social Network Services (groups and individuals). In this paper, we describe an experiment with 48 participants using access controls that include tie strength and tags (separately and simultaneously) together with groups and individuals. We analyze the results using several quantitative and qualitative metrics. We find that users consider these two new attributes useful to define their sharing policies and they prefer to employ access controls that consider tags and tie strength. Specifically, users believe that tie strength improves policy understandability and tags help users define sharing policies faster. However, we also observe that when users employ these two attributes they tend to make more mistakes in terms of the resulting sharing policy. We hypothesize that this could be caused by the lack of experience employing tie strength and tags in an access control.

5.1 Introduction

Millions of photos are shared everyday on Social Network Services (SNSs) such as Facebook, Instagram, and Google+. SNSs enable users to set sharing policies to control how photos are shared with other users. These sharing policies are defined as rules that specify who can access a photo. The *who* can be set employing two elements: groups and individual identifiers. Groups can be defined by each user (e.g., Google+ circles or Facebook friend lists) or can be predefined by the SNS (e.g., friends of friends). These two elements (groups and individual identifiers) allow definitions such as ”share this photo with everyone but Bob”.

As defined by Nissenbaum [109] in the contextual integrity framework, personal information moves from one context to another based on norms of information flow.

When one of such norms is violated, a privacy breach occurs. When SNS users define their sharing policies, they are basically defining their norms of information flow. However, as a number of studies have shown, it is difficult for SNS users to set their sharing policies appropriately using access controls with only groups and individuals as attributes [94, 147]. Consequently, current access controls may not be appropriate to guarantee that the personal information of SNS users flows only through contexts that they consider appropriate.

We identify two key elements of the contextual integrity framework that current access controls for SNS are lacking: tie strength and content type. Nissenbaum points out that distinctive relationships, for example individual to spouse, boss, friend, colleague, and so on, are partially defined by distinctive patterns of information sharing. This was already studied by Granovetter in 1973 [56]. On social media, research works have looked into how people share with contacts of varying tie strength (e.g.,[156]). The results presented in these works show that the tie strength of a relationship is an important factor that users consider when deciding whether disclosing information is appropriated or not.

As explained by Nissenbaum, the definition of norms of information flow is based on the type of the information. For example, people usually do not share health and medical information with strangers. Since we focus our research on photo sharing, we propose the use of tags to define the content of the photo. Currently, SNSs do not allow users to define sharing policies based on the content of the information. However, some photo sharing sites, such as Flickr, enable users to assign tags to their photos so they can be found and classified easily.

Our key objective in this paper is to evaluate the actual impact that tags and tie strength would have on access controls for SNSs. We design a study where 48 Facebook users employ different access controls to express their privacy preferences on a selection of 30 photos of their own. Employing the data collected in the study, we evaluate the impact of tie strength and tags on access controls for photo sharing. This evaluation is based on quantitative measures of their performance, as well as a qualitative analysis of participants' self-reported perceptions.

Contributions

Our contribution is threefold:

1. We evaluate four attributes for access controls: groups, individuals, tie strength, and tags.
2. To evaluate the attributes and the access controls that use them, we propose novel quantitative metrics.
3. We present a qualitative evaluation of tie strength and tags as attributes for access controls. Additionally, employing thematic analysis on the qualitative data, we identify three characteristics that users value from access controls.

5.2 Related Work

A number of papers propose formal access control models for SNSs [18, 22, 46]. Some of these models consider the inclusion of tie strength as an attribute, but none of them considers tags. The authors of these works only focused on the expressiveness of the proposed models. Our work investigates the viability of including these attributes and how they affect access controls.

Access controls for photos that take tags into account have been prototyped by Yeung et al. [163]. Besides, Hart et al. [61] proposed a mechanism to manage privacy for blogs based on tags. Their mechanism enables users to define groups manually or to group potential viewers by attributes that they all share (e.g., workplace or same school). The main focus of their study is to compare basic sharing policy mechanisms for blogs with a tag-based approach. The authors did not use real data from the participants, instead, they asked participants to manage sharing policies of artificial users in hypothetical scenarios. Thus, they do not examine users' actual preferences. Their results show that an approach that uses tags is more appealing to users than one that does not. Klemperer et al. [83] evaluated the usability of an access control based exclusively on tags. The authors aimed at evaluating whether tags can be used to organize pictures and define their privacy at the same time. We, however, study the viability of including tags as a new attribute for access controls along with other attributes that exist already (groups and individual identifiers) and a new one, tie strength.

Besides access control complexity [84], setting privacy preferences can be a tiresome task. To help users cope with this task, a number of privacy recommender tools have been developed [2, 24, 39, 77, 125, 134]. These tools alleviate the burden of privacy configuration on SNSs. The tools proposed in these works aim at facilitating privacy configuration on current access controls. Therefore, they take into account the same attributes of these access controls. We believe that an SNS that offers a combination of access controls that employ attributes that help users understand the implication of privacy settings and privacy recommenders will enhance users' experience.

5.3 Method

We designed a laboratory study during which participants employed different access controls to specify sharing policies for photos of their own. Observing and analyzing how the participants defined their sharing policies, we can evaluate the performance of tie strength and tags as attributes for access controls. Moreover, participants provided their personal impressions about these two attributes. This qualitative data helps us to test whether users perceive the inclusion of these new attributes as beneficial for managing their privacy on social media.

Our investigation is based on real data that was collected from Facebook users. We specifically chose Facebook users for two reasons: (i) Facebook users are accustomed to share information on the Internet and to employ sharing policies to protect it; and

(ii) Facebook is one of the most successful SNSs, thus, there are many potential participants.

To recruit participants, we published the study on a number of social media sites. We relied on the viral properties of social media to attract diverse participants out of the academic environment. Therefore, as people signed up for the study, we asked them to share their participation on their social media accounts. To reward participants for their time and effort, we gave them a 10Euro gift voucher.

The information required from participants was varied, thus, we divided the study in six steps. In each step, the participants provided a different type of information. The steps were: (i) collection of social data, (ii) choosing photos, (iii) tagging photos, (iv) definition of sharing policies, (v) correction of sharing policies, and (vi) post-survey questionnaire. These steps are explained with further detail in following sections. The average time to complete the study was around one hour and a half. Since completing the study in one session could be tiresome, it was possible to stop and resume it at any time.

During the study, participants provided personal photos and we collected sensitive data from their Facebook accounts. We informed them beforehand that their collected data would be only used for academic purposes. Moreover, we committed to anonymize their data as soon as they finished the study. Likewise, participants were informed that the photos employed during the study would not be seen by any member of our group at any time and that we would delete them as soon as they finished the study. We obtained approval for our study from the ethics committee at Universitat Politecnica de Valencia; approval was made in writing.

5.3.1 Demographics

The sample consisted of 23 women and 25 men (48 total). Regarding ages, 52% were between 18 and 24, 23% between 25 and 29, 21% between 30 and 39, and 4% had ages between 40 and 49. For education level, the majority of them had a college degree (23), 20 of participants had a high school degree, 3 of them had a PhD, and two participants had a primary school degree. Finally, 72% of the participants were students and the other 28% were working.

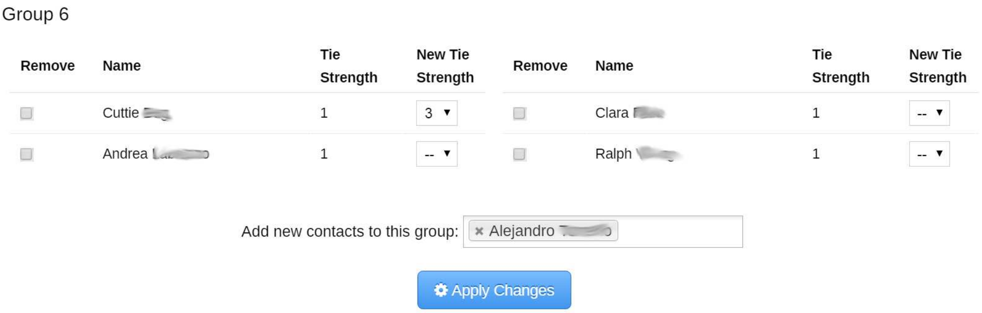
5.3.2 Step 1: Collection of Social Data

During the evaluation, we employ data collected from participants' Facebook accounts. Specifically, their relationship information (i.e., how they group their contacts on Facebook and tie strength values). To collect this information, we used BFF [41]. This is a Facebook application that helps users organize their relationships. BFF automatizes the process of friend grouping and tie strength definition. BFF represents tie strength on a Likert scale 1-5 (1 = minimum, 5 = maximum). Thus, we employed the same level of granularity to represent tie strength levels.

BFF was only employed as a way to collect data from Facebook and speed up the process of defining the characteristics of the social relationships of the participants. Hence, its results could be modified as the participants considered necessary. It is worth noting that, like in Facebook, contact groups were not exclusive, thus, one single contact could be included in several groups if the participant found it appropriate.

Fig. 5.1 shows the user interface that presents the results obtained by BFF and offers the possibility to the user to adjust them. In the example shown in the Fig. 5.1, the user is adding a new member to a group and changing the tie strength of a contact (*Cuttie*) from 1 to 3.

As shown in [83], when people employ tags to specifically manage their privacy, they behave differently to when they employ the tags as a means to organize their photos. Likewise, it can be expected that users employ other attributes, such as tie strength, differently if they are told beforehand that they will be using those attributes for privacy management. Since we wanted to observe how users manage their sharing policies employing the different attributes, participants were instructed to group their

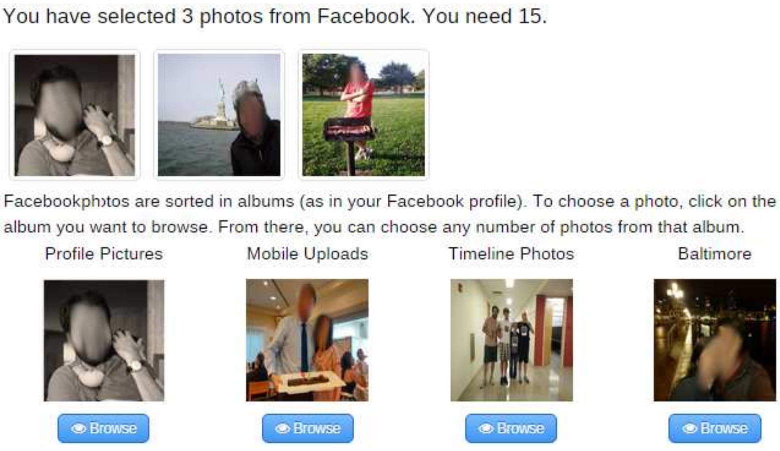


Group 6

Remove	Name	Tie Strength	New Tie Strength	Remove	Name	Tie Strength	New Tie Strength
□	Cuttie [redacted]	1	3 ▾	□	Clara [redacted]	1	-- ▾
□	Andrea [redacted]	1	-- ▾	□	Ralph [redacted]	1	-- ▾

Add new contacts to this group: × Alejandro [redacted]

Figure 5.1: Tie strength and group correction interface.



You have selected 3 photos from Facebook. You need 15.

Facebookphɔtos are sorted in albums (as in your Facebook profile). To choose a photo, click on the album you want to browse. From there, you can choose any number of photos from that album.

Figure 5.2: Photo selection interface.

contacts, tag their photos, and assign tie strength values taking into consideration that they would be using that information to define their sharing policies.

5.3.3 Step 2: Choosing Photos

Since we wanted to collect information on how users manage the privacy of their photos in a variety of situations, we asked participants to choose 15 photos from their Facebook accounts and 15 photos that they had not shared on any SNS. The idea behind this was to collect photos that participants did probably not share on any

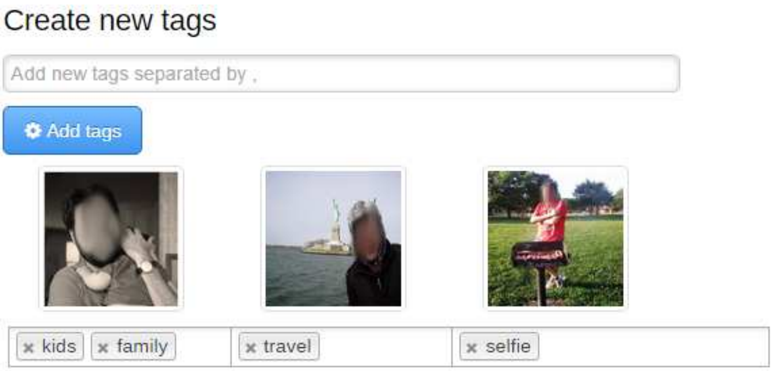


Figure 5.3: Assigning tags to photos.

SNS due to privacy concerns. In this way, we can evaluate new access controls and attributes on photos that users believe current access controls do not protect well. We instructed participants to choose 15 photos from outside Facebook that were suitable to be shared with at least one contact on Facebook, despite they were not shared on any SNS. In this fashion, we avoided participants choosing photos so sensitive that their only appropriate sharing setting was to keep them private. Additionally, we asked participants to choose photos that depicted specific items. We brainstormed a list of 17 items which included elements such as selfie, friends, special day, kids, a moment with your significant other, and food. Participants were not required to cover every item, but we asked them to cover as many as possible. The goal of this requirement was to force participants to choose photos with the widest possible variety. Fig. 5.2 shows the interface used by participants to choose their 30 photos.

5.3.4 Step 3: Tagging Photos

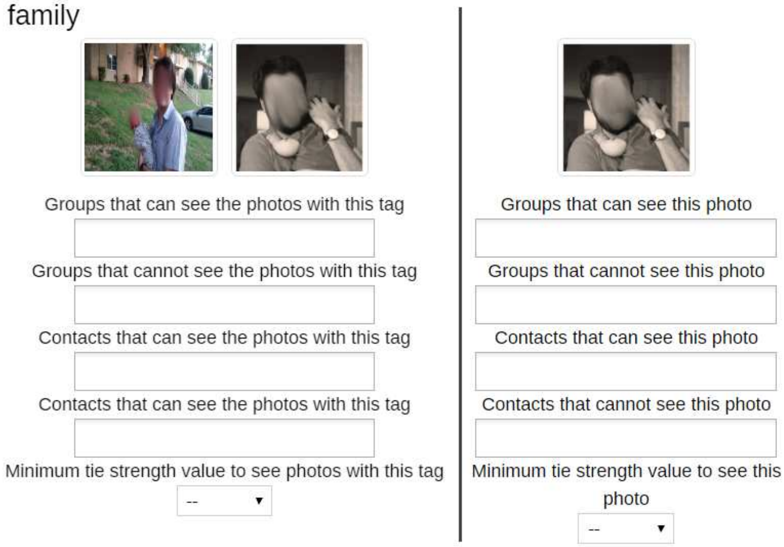
After the participants chose their photos, they had to assign tags to them. The tags were free and each participant defined as many as he or she felt necessary. Each photo had to be classified with at least one tag and a photo could be classified employing many tags. Fig. 5.3 shows the interface the participants employed to assign tags to their photos. In the example figure, the shown photos have been tagged as *family* and *kids*, *travel*, and *selfie* respectively.

5.3.5 Step 4: Definition of Privacy Preferences

In this step, the participants had to specify their privacy preferences employing policies defined with two different access controls. These access controls had different combinations of the following attributes:

- **Tags (*Tag*)**: Labels or categories that users employ to organize their photos by their content (e.g., family, and travel).
- **Groups (*Group*)**: Groups of contacts created by the users.
- **Individual Contacts (*Ind*)**: Individual identifiers of each contact of the user.
- **Tie Strength (*Tie*)**: The level of closeness that users have with each one of their contacts based on a Likert scale from 1 to 5 (1 = weak tie, 5 = strong tie).

Employing these attributes, we defined three different access controls: TagGroupIndTie, GroupIndTie, and TagGroupInd. We name the access controls based on the short names of the attributes that each one employs. For example, GroupIndTie employs groups, individual contacts, and tie strength. We employed these specific three access controls because they enable us to compare, together and separately, tie strength and tags with groups and individuals. The most direct approach to collect data to compare all access controls would have been to ask participants to express their privacy preferences with all access controls. This approach was not feasible as it would have made the study too demanding. Participants would have been required to assign a sharing policy to each photo three times (one for each access control). This may have fatigued them, which, in turn, may have led to sloppy responses. Instead, we asked participants to set their privacy preferences employing only two of the access controls. In this fashion, they could



family

Groups that can see the photos with this tag

Groups that cannot see the photos with this tag

Contacts that can see the photos with this tag

Contacts that can see the photos with this tag

Minimum tie strength value to see photos with this tag

-- ▾

Groups that can see this photo

Groups that cannot see this photo

Contacts that can see this photo

Contacts that cannot see this photo

Minimum tie strength value to see this photo

-- ▾

Figure 5.4: Sharing policy definition (tag view on the left and photo view on the right).

compare the benefits and disadvantages of one combination of attributes over the other.

Fig. 5.4 shows the interface that participants used to define their sharing policies. While using TagGrouIndTie or TagGroupInd, participants employed an interface to define sharing policies that offered two views: photo and tag. Photo view enabled participants to define their sharing policies on a per photo basis. On the other hand, tag view enabled them to define sharing policies on a per tag basis. These two views were accessible through tabs. Fig. 5.4 shows the tag view on the left and the photo view on the right.

Regardless of the access control, sharing policies were defined in terms of two predicates: allow and block. Contacts that are referenced by any attribute in the allow predicate are granted access to the photo (or photos with a specific tag), and vice versa for the block predicate. Following the same approach as mainstream SNSs, block takes precedence over allow. The tie strength attribute worked slightly different. Participants employing either TagGroupIndTie or GroupIndTie could specify a tie strength threshold in their sharing policies. All contacts with a tie strength value equal or higher than the threshold were granted access to the photo or tag, the others were blocked. Finally, by default, no contact had access to any given photo. All this was explained to the participants before they started the study and they were reminded during this step.

We collected the same number of samples for every access control; specifically, each access control was employed 32 times (48 participants × 2 access controls ÷ 3 access controls). Moreover, the order of how the access controls were presented to the participants was random. This measure was taken to counteract any ordering bias, e.g., participants getting used to the interface of the survey application, thus, finding themselves more comfortable with the second access control. Finally, we also assigned randomly the order in which tabs for tags and photo views where shown. In this way, we also removed any bias generated from showing any of those views first.

5.3.6 Step 5: Correcting Sharing Policies

After participants defined their sharing policies employing both access controls, the next step was to correct errors on sharing policies. Participants were using new access controls with attributes that they had not used before, thus, some mistakes could be expected. Additionally, participants defined their sharing policies employing two different access controls, divergences between sharing policies for the same given photo were likely to occur.

Since participants employed two access controls, each photo had two associated sharing policies. In this step, each photo was shown in conjunction with the list of contacts allowed and not allowed to see it according to each policy. In this fashion, participants could review thoroughly the sharing policies they defined in the previous step. When the two sharing policies for the same photo (one for each access control) were the same, participants could choose to mark them as correct or mark them as incorrect and modify them. On the other hand, if the policies were different, participants could either choose one as correct or none. If they did not chose anyone, they had to modify one of the two.

5.3.7 Step 6: Post-Survey Questionnaire

In the last step of the study, participants responded to a survey about the access control they liked the most. Participants were required to articulate their decision as much as possible. They also reported some information about privacy practices and how they use Facebook. Specifically, they reported the following data: Facebook account age, frequency of use of Facebook, number of photos on Facebook, level of privacy concern while using social media, whether they changed the default sharing policies of Facebook or not, whether they created one friend list on Facebook or not, level of knowledge about Facebook's use of personal data policy, and how often they specify a sharing policy while sharing a photo on Facebook.

5.4 Evaluation

In this section, we describe our hypotheses and evaluate them.

5.4.1 Hypotheses

The hypotheses tested in this paper are:

- *H-Useful-Tie-Strength*: Users find tie strength useful to define sharing policies.
- *H-Useful-Tags*: Users find tags useful to define sharing policies.
- *H-Performance-Tie-Strength*: Users can express accurately their privacy preferences when they employ sharing policies that use tie strength.

- *H-Performance-Tags*: Users can express accurately their privacy preferences when they employ sharing policies that use tags.
- *H-Preference-Tie-Strength*: Users prefer to employ access controls that include tie strength.
- *H-Preference-Tags*: Users prefer to employ access controls that include tags.

5.4.2 *H-Useful-**

To evaluate these hypotheses, we measure the usage of each attribute in the policies defined by the participants. We count as one use of an attribute to specify one or more values for that attribute. For example, the policy *Allow*{Group(friends), Alice}, *Block*{Group(colleagues)} uses the attributes group and individual. The usage of an attribute is the average number of times that the given attribute was employed by a participant to define a sharing policy employing an access control that considers that attribute. For instance, if a participant using GroupIndTie employed group to define the sharing policies of 15 of his 30 photos, the usage of group for that participant would be 50%. Our intuition is that the more a user employs an attribute, the more he/she feels it is useful to define sharing policies.

Fig. 5.5 is a boxplot that shows the average usage of each attribute per participant. The line inside the boxplot indicates the median. The tops and bottoms of each ”box” are the 25th and 75th percentiles of the samples, respectively. Whiskers are drawn from the ends of the interquartile ranges to the furthest observations that are not considered outliers. An outlier (represented as a cross) is a value that is more than 1.5 times the interquartile range away from the top or bottom of the box. Finally, the notches of the boxes display the variability of the median between samples. Boxes whose notches do not overlap have different medians at the 5% significance level.

On the one hand, tag, group, and tie are intensively used and their differences in use are not statistically significant; however, tie strength presents a higher variance. On

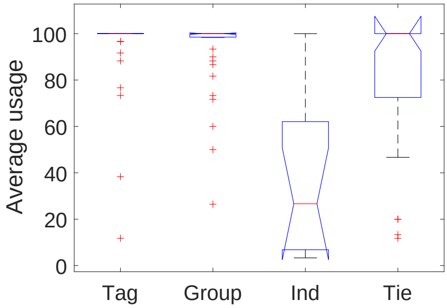


Figure 5.5: Average attribute usage per participant.

the other hand, individual is much less used than the other three attributes.

5.4.3 *H-Performance-**

To evaluate *H-Performance-** hypotheses, we build *regression* models with multiple predictors and one response variable. To build the models, we consider each sharing policy defined by the participants as one sample. Since each participant had to define the sharing policy of 30 photos employing two access controls, the dataset contains 2880 samples (48 participants × 30 photos × 2 access controls). The features of the samples are dummy variables that represent whether an attribute was used or not in the policy. For example, the policy *Allow*{Group(friends, family), Alice}, *Block*{Group(colleagues)} uses the attribute group and the attribute individual identifier. Therefore, in this example, the variables *group used* and *individual used* equal 1, while *tag used* and *tie strength used* equal 0. These dummy variables are used as predictors in the model. When analyzing the output of the model, we focus on *coefficients* and their statistical significance. These coefficients help us understand each feature's relative influence on the response variable.

Since each participant defined several sharing policies, each participant's personal experience with access controls may affect the estimated coefficients of the models. Thus, we employ mixed modeling [121] to create all the regression models.

Mixed models offer the possibility of grouping samples of the data hierarchically. Specifically, we group sharing policies by participant by employing the participant ID as a *random effect*. In this way, the participant ID captures the variability introduced by personal experiences with access controls in the responses. Introducing the ID enables us to obtain estimated coefficients for the other predictors (*fixed effects*) in a way that is less affected by the variability in the participants' responses. Finally, to show the significance of the coefficients, we employ * and ** that indicate p<0.05 and p<0.01, respectively.

The first model is a *logistic regression model* where the response variable is categorical and indicates whether the policy is correct or not. Since participants had to review and mark as correct or incorrect each policy, the computation of this response variable is straightforward.

In logistic regression models, the coefficients are expressed in log-odds units. The coefficients show the effect of a predictor on the log odds of the sharing policy being in a given category ($Y = 1$) versus being in the reference category ($Y = 0$); that is, the odds of Y being a given category increase by a factor of e^{b_i} per unit change in X_i. The equation used by the logistic models is

$$P(Y = 1) = \frac{1}{1 + e^{-b_0 - \sum b_i X_i}} \tag{5.1}$$

where b_i is the coefficient of predictor X_i.

Table 5.1 shows the output of the multinomial logistic regression. Our intuition is that an attribute with a negative coefficient indicates a disconnection between how the user believes the attribute works in terms of blocking and granting access to contacts and how it actually does. Specifically, employing tags increases the risk of defining an incorrect policy, while employing groups and individual identifiers reduces it. The influence of tie strength on the probability of a policy being correct/incorrect is not significant.

The logistic regression model assumes that each predictor is independent. However,

Table 5.1: Coefficients for policy correctness regression model.

	Correct = True
Tag Used = True	-0.528^{**}
Group Used = True	0.745^{**}
Ind Used = True	0.402
Tie Used = True	-0.079

Table 5.2: Coefficients for policy correctness regression model employing access controls as predictors.

	Correct = True
Access Control = TagGroupIndTie	-0.867^{**}
Access Control = TagGroupInd	-0.379^{*}

in the study, participants defined policies combining different attributes, thus, this assumption may not be true. To study how the combination of attributes affects the probability of a policy being correct, we create a new logistic regression model with the access control employed to define the policy as the predictor variable. Table 5.2 shows the coefficients of each access control (GroupIndTie is the reference category, hence, it is not shown in the Table). The coefficients indicate that a policy defined with TagGroupIndTie or TagGroupInd has a higher risk of being incorrect than one defined with GroupIndTie.

To explore further the relationship between policy correctness and attribute combinations, we build a C4.5 decision tree [118] employing Weka implementation [49], which achieves a 74.4% accuracy. Fig 5.6 shows the resulting decision tree, where each leaf represents a class (i.e., correct or incorrect) and shows its name and the percentage of instances reaching that leaf. Our objective in training the decision tree is to interpret the paths in the tree. Our intuition is that the path from the root to a leaf in the tree describes the employed combination of attributes that leads to a sharing policy being correct or incorrect.

We observe that the way tie strength is combined with other attributes is crucial

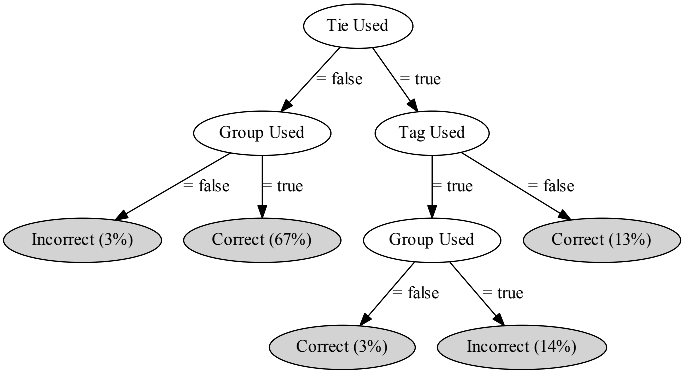


Figure 5.6: A decision tree with the correctness of policies as decision targets (leaf nodes; shaded).

for the correctness of a policy (tie strength appears in the root node of the tree). On the one hand, if tie strength is combined with groups and tags simultaneously, incorrect policies can be expected. On the other hand, if tie strength is not employed, or employed without tags, the chances of the policy being correct increase.

The models created with the correctness of the policy as the response variable provide a measure of the influence that attributes have on defining an incorrect policy. However, they do not measure the influence of the attributes on the seriousness of the privacy breach produced by the incorrect policy. For example, the privacy breach produced by a policy that allows incorrectly access to a single contact is likely less serious than the breach produced by a policy that incorrectly grants access to dozens of contacts. To this aim, we calculate the mutual information between each policy marked as incorrect by the participants and its corresponding corrected one. The mutual information is in range [0,1], it equals 0 when both policies (the incorrect and the correct) are completely disjoint, and it is close to 1 when the differences are minimal. Since the mutual information is a continuous variable, we build a *multiple linear regression model* with the mutual information as the response variable. Linear regression models use the general linear equation

$$Y = b_0 + \sum b_i X_i \tag{5.2}$$

Table 5.3: Coefficients for mutual information regression model.

Predictor	b
Tag Used = True	0.059
Group Used = True	0.011
Ind Used = True	0.142**
Tie Used = True	0.058*

Table 5.4: Coefficients for mutual information regression model employing access control type as predictor.

Predictor	b
Access Control = TagGroupIndTie	0.027
Access Control = TagGroupInd	-0.008

where Y is a continuous response variable and b_i is the coefficient of predictor X_i. As in the logistic regression models we employ mixed modeling to reduce the influence of personal experience with access controls on the coefficients. Table 5.3 shows the coefficients yielded by the linear regression model. Analyzing the model output, high coefficients indicate that the attribute contributes to reduce the privacy breach caused by an incorrect policy. Specifically, individual identifiers is the attribute that reduces privacy breaches the most, followed by tie strength. Using groups and tags is not statistically significant.

We now build a linear regression model using the type of the access control employed to define the policy as the predictor and the mutual information as the response variable. Table 5.4 shows the coefficients yielded by the model. In this case, no coefficient is statistically significant. Thus, there is no evident correlation between the access control employed and the mutual information between an incorrect policy and its corresponding correct one.

The last analysis to test *H-Performance-** hypotheses is to measure the influence of attributes and their combinations on policy redundancies. Sharing policies can include several times the same contact in the allow or block predicates. For

Table 5.5: Coefficients for positive redundancies model.

Predictor	b
Tag Used = True	0.408**
Group Used = True	0.223**
Ind Used = True	0.082**
Tie Used = True	0.356**

Table 5.6: Coefficients for negative redundancies model.

Predictor	b
Tag Used = True	0.098**
Group Used = True	0.1**
Ind Used = True	0.057**
Tie Used = True	0.16**

example, a user can block a contact from seeing a photo by blocking the group the contact belongs to and by setting a tie strength threshold higher than the contact's. Redundancies measure the number of times each contact was included several times in the block or allow predicate. Values are relative to the number of contacts a given participant has. Redundancies can be positive or negative depending on the contact being allowed or blocked several times. High numbers of redundancies can lead to verbose policies, which is detrimental to policy definition and maintenance. The number of redundancies is a continuous variable, thus, in this analysis we employ linear regression models. As in the previous models, we employ mixed modeling. Table 5.5 shows the coefficients yielded by the model. The more attributes are used in a sharing policy, the greater the probability of redundancies, thus, all coefficients are significant. Furthermore, the attributes that cause positive redundancies the most are tags and tie strength.

Now we create a linear regression model for negative redundancies. Table 5.6 shows the output of the model. Again, all coefficients are significant. However, this time, the attributes that cause negative redundancies the most are tie strength and groups.

Table 5.7: Coefficients for positive redundancies model employing access control type as predictor.

Predictor	b
Access Control = TagGroupIndTie	0.469**
Access Control = TagGroupInd	0.144**

Table 5.8: Coefficients for negative redundancies model employing access control type as predictor.

Predictor	b
Access Control = TagGroupIndTie	0.2**
Access Control = TagGroupInd	0.029

Finally, we create two linear regression models employing the access control type as the predictor variable. Tables 5.7 and 5.8 show the coefficients for positive and negative redundancies respectively. According to the coefficients yielded by the model, when employing TagGroupIndTie, users have greater chances of creating policies with redundancies.

5.4.4 *H-Preference-**

To evaluate *H-Preference-** hypotheses we employ participant's responses about what access control they liked the most. TagGroupIndTie was chosen by 65.63% of those who tried it, participants who tried GroupIndTie chose it as the best 53.12% of the time, and the least preferred was TagGroupInd, which was chosen 31.25% of the time. Since each participant was assigned two access controls in the study, they chose their preferred access control over the other. Table 5.9 shows the preferred percentage in pairs. Each row in the table shows the percentage of times that the access control was chosen as the best over the other in the column.

We employ a multinomial logistic regression to find what demographic factors influence the most on the preferred access control. Table 5.10 shows the coefficients yielded by the regression employing the preferred access control as the response

Table 5.9: Percentage of times each access control was preferred over each other.

	TagGroupIndTie	**TagGroupInd**	**GroupIndTie**
TagGroupIndTie	×	68.75%	62.5%
TagGroupInd	31.25%	×	31.25%
GroupIndTie	37.5%	68.75%	×

Table 5.10: Regression coefficients for demographic variables.

	TagGroupIndTie	**TagGroupInd**
Gender = Woman	−1.501	−0.467
Age	−0.877	−1.242
Education level	2.772	2.775
Facebook experience	−19.94**	−20.675**
Facebook frequency of use	0.309	0.05
# friends on Facebook	0.644	0.891
# photos on Facebook	2.022*	1.742
Privacy concerns on social media	2.815*	2.566
Default sharing policies changed = Yes	41.512**	−50.729**
Friend lists created = Yes	0.134	0.465
Knowledge of Facebook data policy	−1.602	−1.478
Frequency of sharing policy setting	−1.104	−1.391

variable, and using GroupIndTie as the reference category. Most of the variables are not statistically significant or have low coefficients. There are two exceptions: Facebook experience and whether the user has changed her default sharing policies or not. On the one hand, experienced users are inclined to choose GroupIndTie. On the other hand, users that have changed their default sharing policies on their Facebook accounts prefer TagGroupIndTie over GroupIndTie, and GroupIndTie over TagGroupInd.

To further explore the preferences of the participants, we classify the given explanations based on the explicit mention of tie strength and tag attributes. Tie strength was the explicit reason why 31.25% of the participants chose an access control over the other. The majority of these participants highlighted the granularity that tie strength provides when classifying contacts. For example, participant P12 said *"It is easier to use. In a given group you may have contacts with high and low proximity, which would require to split that group in smaller subgroups to create appropriate sharing policies"*. Some participants also indicate that tie strength makes easier to manage new contacts that have not been assigned to a group yet (P28 *"I think that tie strength is important if you do not have a good set of groups or if you just added somebody and you have not assigned a group to them yet"*).

Participants mentioned explicitly that tags were the reason to chose one access control over the other 18.75% of the time. The majority of these selections were justified by the simplification that tags provide when defining sharing policies. For example, P1 explained: "*In the first access control, photo categories speed up the whole process. It requires careful organization, though.*". Some participants also valued the tag view and its organizative functionality (P21 "*I prefer this access control because it offers two perspectives. Out of these two perspectives, I like tag the most because it enables me to: 1) organize my photos by category, and 2) since I usually assign the same [privacy] preferences to photos in the same category, it is faster.*)"

We find that 9.25% of the participants disliked tags. These participants considered tags only as a mechanism to organize pictures and failed to see how this attribute could help them to manage their privacy. Finally, we find only one participant that disliked tie strength. This participant pointed out that assigning tie strength values to every contact was a time consuming task: "*In practice, we do not have that much time to consider who can and who cannot see your photos; it is too time consuming. I think grouping contacts is enough, there is no need for tie strength.*"

5.5 Discussion

The results show that tags and tie strength are extensively employed by users. Therefore, users seem to find that these two attributes are useful to define their sharing policies. Furthermore, analyzing the self reported preferences of the participants, we find that access controls that consider the attribute tie strength are preferred by users. It is worth noting that 30% of participants mentioned tie strength as the reason why they chose one access control over the other. Finally, we find that users who are experienced on Facebook and have modified, at least, the default privacy settings of Facebook prefer TagGroupIndTie and GroupIndTie over TagGroupInd.

Analyzing individually the effects of tie strength and tags on the correctness of sharing policies, we find that tag is the attribute that increases the risk of defining an incorrect policy the most. A study on how combinations of attributes impact policy correctness shows that users create the majority of incorrect policies when combining tie strength and tags. This points out that when users create complex policies, they find difficult to foresee the implications of the combinations of these two attributes on the potential viewers. However, we also find that when an incorrect policy is created, tie strength can help to reduce the seriousness of the privacy breach. We posit that, since every contact has a tie strength value assigned, this attribute functions as a safe default that prevents acquaintances from accessing sensitive photos.

Finally, we find that tie strength and tags do not increase the ability of users to define succinct policies. Although, group has a similar effect on the verbosity of sharing policies.

In general, the results show a disconnection between the performance of the attributes and the usage and preferences of the users. On the one hand, the metrics employed to evaluate attributes and access controls indicate that tie strength and tags do not perform as well as previous literature had anticipated. On the other hand, users employ intensively these two attributes when defining sharing policies and they prefer an access control that offers them. Specifically, users that are experienced on Facebook and its privacy controls.

5.5.1 Thematic Analysis

To discover the reasons why users prefer access controls with tags and tie strength, we perform a thematic analysis on the explanations that the participants gave when choosing their preferred access control. This form of analysis goes beyond simply counting phrases or words in a text and moves on to identifying implicit and explicit ideas within the data. To that aim, we define three categories that classify the impressions of the participants:

- **Understandability**: Participants valued how easy was to comprehend who will be allowed to see each photo and who will not. An example of this class is: ”*This access control is easier to manage. There may be people in the same group that has different degrees of closeness, so you need to be careful and pick them individually. However, employing tie strength, you know, for sure, that only very close people can see the photo.*”.
- **Granularity**: A number of participants perceived that some access controls offered finer levels of granularity that enabled them to carefully tailor who could view their personal information. An example of this class is: ”*I prefer the second access control because I can make a better selection of who can and who cannot see my photos. It allows me, not only to choose a group, but also to include some people from another group or remove some from the selected group so they cannot see the photo.*”
- **Speed of configuration**: While some participants preferred to have fine grained access controls, other were more satisfied with access controls that can be set quickly. An example is: ”*Access control two requires too much time. It is faster to create small groups and categorize the photos.*”

Overall, the classification of explanations was as follows: understandability (26.31%), granularity (28.95%), speed of configuration (44.74%). With these categories, 75% of the explanations were classified. Those that could not be classified

Table 5.11: Spread of explanation classes across access controls.

	TagGroupIndTie	TagGroupInd	GroupIndTie
Understandability	70%	20%	10%
Granularity	45.46%	18.18%	36.36%
Speed	23.53%	47.06%	29.41%

were either too vague or the participant failed to clearly express her perception. Categories were not exclusive, thus, an explanation could be classified into two or more classes.

Table 5.11 shows the explanation class spread for all the access controls. Interpreting the results, participants considered that it is easier to understand the consequences of a sharing policy defined with an access control that combines tie strength and tags. They also considered that tie strength offers high levels of granularity when defining sharing policies. Furthermore, participants valued the speed that tags offer; enabling them to set sharing policies faster.

5.5.2 Practical Implications

According to the results, users value positively the addition of tie strength and tags, however, the results also indicate that users will require assistance when using these attributes. Furthermore, adding new attributes to an access control can increase the effort that users have to dedicate to set their privacy preferences. However, there are a number of automated approaches that can alleviate these problems. We envision an access control that offers attributes that users can employ to effectively define their privacy preferences while being easy and quick to set up. To illustrate how a user could deal with these new attributes, imagine an SNS which offers the following functionalities:

- **Automatic contact grouping and tie strength computation**: Considering that the average number of contacts Facebook users have is around 160 [37], grouping contacts and specifying a tie strength value for each one can be a daunting task. Therefore, SNSs should offer tools that alleviate this task. A number of research works propose tools that can help users group their contacts [2, 39]. Similarly, tools such as BFF [41] can predict tie strength with an acceptable accuracy.

- **Automatic photo tag inference**: It can be expected that each user shares habitually the same types of photos. A tool such as the one proposed by Kucuktunc[86] can learn tags from previous photos shared by a user and tag new photos accordingly. Therefore, users would only need to specify tags for a few initial photos. New photos would be automatically tagged, speeding up the process of defining sharing policies employing tags.

- **Personalization**: Current access controls for SNS are static; they do not adapt to users. However, as the qualitative evaluation shows, experience with social media influences the choice of access control. We posit that users should be able to configure the attributes employed by the access control. In this way, as users get accustomed to how photo sharing works on SNS and its implications, they could adapt the access control to meet their specific needs.

- **Sharing policy assistant**: The results obtained by tie strength and tags in the correctness test and by all attributes in the redundancy test point out that users require assistance when defining sharing policies. A visualization tool similar to the one proposed by Lipford et al. [95] can help user understand the implications of a sharing policy. Similarly, a misconfiguration detection utility, such as [75], could indicate to users what sharing policies have more chances of being incorrect. Finally, to reduce the redundancy of policies, the SNS could offer a policy simplification tool such as the one presented in [5].

5.5.3 Limitations

There is a number of limitations to our study design. First, our results are limited by the participants we recruited and the photos they provided. As shown in

Demographics, the sample of participants was skewed towards young people with a high degree of education.

We asked participants to provide photos that fulfilled a list of 17 requirements. However, the list may not be comprehensive enough and not cover every type of photo that users normally upload to an SNS. Therefore, the generalizability of the photos provided, which may affect the results obtained by the attributes and access controls, was also limited to what participants considered appropriate.

Other limitations concern scale. We cannot comment on whether all access controls and the two new attributes would remain viable when dealing with thousands of photos and hundreds of contacts of varying types (friends, family members, and acquaintances). Further work should be done on this regard.

Finally, our study is cross-sectional (one time). Our study does not provide data to understand how users deal with access controls and their attributes over a period of time. This might be the reason why, although users like tie strength and tags, these two attributes did not perform as well as the more known attributes group and individual. It is possible that for users to obtain the full benefit from these two new attributes, they need to refine iteratively tie strength values and tags as they learn the effects of these attributes on their sharing policies.

5.6 Conclusions

Social media has become one of the most common and useful ways of sharing photos. Nonetheless, unsatisfactory privacy management on SNS lead to privacy leaks that may drive users away from employing all the functionalities offered by SNSs. In order to improve privacy management on SNSs, developers have to offer access controls that are easy to understand and make users feel that they are in control of how their information is disseminated. Related literature had identified tie strength and tags as attributes with potential to improve access controls. In this paper, we evaluate their viability.

An analysis on the preferences and use of each attribute indicates that users value positively tie strength and tags and find them useful to define sharing policies. Moreover, a qualitative analysis informs about what features users find important in an access control. Based on this, users value the granularity and understandability that tie strength offers and how tags can speed up the privacy configuration process. Nonetheless, when users employ these two new attributes, they tend to make more mistakes in terms of policy correctness. Although, tie strength can help reduce the damage that a privacy breach may cause. We hypothesize that results obtained by tie strength and tags in terms of policy correctness and redundancy can point to a lack of experience dealing with these attributes in an access control. Therefore, SNS developers willing to include these attributes should consider first the addition of tools and mechanisms that help users understand the implications of tie strength and tags on their privacy. At least, until users become accustomed to their use. Future work should test whether users educated about how to use tie strength and tags make fewer mistakes than those who use them for the first time.

CHAPTER

6

Sharing Policies in Multiuser Privacy Scenarios: Incorporating Context, Preferences, and Arguments in Decision Making

AUTHORS:
RICARD L. FOGUES[⋆], PRADEEP MURUKANNAIAH[§], JOSE M. SUCH[†], AND MUNINDAR SINGH[♠]
rilopez@dsic.upv.es, pkmvse@rit.edu, jose.such@kcl.ac.uk, mpsingh@ncsu.edu
[⋆]DEPARTAMENTO DE SISTEMAS INFORMÁTICOS Y COMPUTACIÓN
UNIVERSIDAD POLITÉCNICA DE VALENCIA, SPAIN
[§]ROCHESTER INSTITUTE OF TECHNOLOGY, NY, USA
[†]KING'S COLLEGE DEPARTMENT OF INFORMATICS LONDON, UK
[♠]DEPARTMENT OF COMPUTER SCIENCE AT NC STATE, NC, USA

Abstract

Social network services enable users to conveniently share personal information. Often, the information shared concerns other people, especially other members of the social network service. In such situations, two or more people can have conflicting privacy preferences; thus, an appropriate sharing policy may not be apparent. We identify such situations as *multiuser privacy scenarios*. Current approaches propose finding a sharing policy through preference aggregation. However, studies suggest that users feel more confident in their decisions regarding sharing when they know the reasons behind each other's preferences. The goals of this paper are (1) understanding how people decide the appropriate sharing policy in multiuser scenarios where arguments are employed, and (2) developing a computational model to predict an appropriate sharing policy for a given scenario. We report on a study that involved a survey of 988 Amazon MTurk users about a variety of multiuser scenarios and the optimal sharing policy for each scenario. Our evaluation of the participants' responses reveals that contextual factors, user preferences, and arguments influence the optimal sharing policy in a multiuser scenario. We develop and evaluate an inference model that predicts the optimal sharing policy given the three types of features. We analyze the predictions of our inference model to uncover potential scenario types that lead to incorrect predictions, and to enhance our understanding of when multiuser scenarios are more or less prone to dispute.

6.1 Introduction

A social network service (SNS) enables users to maintain social relationships via online interactions. SNS users share information with each other as they interact. Often, the information shared on an SNS involves more than one user. A natural example is a picture or video showing a group of people. Many SNSs enable users to connect the information they upload to other users so that the connected users can be notified of the uploaded information. Since the information shared varies depending on the SNS, these connections can take different forms, e.g., tags on a picture uploaded to Instagram or mentions in a tweet. Suppose Alice uploads a picture from last weekend's party in which she appears together with her friend Bob, and tags Bob in the picture. When these connections are created, the other users are linked to the uploaded information. Usually, a connection implies that the profile of the user can be accessed from the information or some personal information is shown in conjunction with the uploaded data. Although connections between information and users are widely employed by SNSs, they can pose a privacy threat. For example, Bob may find that the picture Alice uploaded is sensitive. However, Bob has no control over uploading that picture, and Alice's action can threaten Bob's privacy by revealing information about him from one setting or context into another. We identify a situation such as this as a *multiuser privacy scenario* or, for brevity, *multiuser scenario*.

Currently, SNSs do not provide mechanisms to handle multiuser scenarios [42]. Thus, a user who did not upload a piece of information concerning him must deal with the privacy settings chosen by the uploader; at best, the user can remove the connection that links him to the shared information, but the information itself remains nevertheless. An ideal solution in a multiuser scenario is to respect each user's privacy. However, often such a solution may not be viable since the preferences of the users involved may conflict. For example, suppose Alice would like to share a picture in which Bob and she appear along with her friend Charlie. However, Bob would like to share this picture only with his common friends and he does not know Charlie. Here, no solution completely respects both Alice's and Bob's preferences.

Several researchers, e.g., [6, 87, 159, 146], have identified the lack of decision-support systems that help users resolve multiuser privacy conflicts as one of the biggest gaps in privacy management in social media. The main challenge that such decision-support systems would address is proposing an *optimal* solution: a solution most likely to be accepted by all those involved in the multiuser scenario.

Although an optimal solution may not exist for each multiuser scenario, identifying one, when it exists, can minimize the burden on the users of having to resolve the conflict manually.

Other researchers (see [133, 20, 71, 143]) have proposed methods to automatically determine solutions to conflicts based on users' privacy preferences. These methods suffer from two main limitations. One, either they aggregate preferences in the same way regardless of the context or consider only a few, predetermined situations among the large number of potential situations. Two, they do not consider the reasons behind users' preferences. However, evidence from self-reported data [87, 159] suggests that users entertain the explanations provided by others and that the optimal solution may depend upon the particular context and reasons behind users' preferences. Following this idea, we empirically study three types of factors that potentially influence a privacy decision: the scenario's *context*, users' *preferences*, and their *arguments* about those preferences. An argument is a justification a user employs to convince the others involved that the user's expectations are reasonable and should be taken into account for making a decision.

Our key objective in this paper is to build an argumentation-based model that accurately represents a multiuser scenario. To this end, we (1) identify important factors that potentially influence the inference of sharing decisions; (2) evaluate the relative importance of these factors in inferring an optimal solution; and (3) develop a computational model that predicts an optimal solution for a given multiuser scenario.

We design a study where human participants are asked to choose what they think is the most appropriate sharing policy in a multiuser scenario we specified. Combining different values of the three types of factors we identified (context, preferences, and arguments), we generate 2,160 scenarios. Considering the sheer number of participants required, we conducted our study on Amazon Mechanical Turk (MTurk), collecting responses from 988 unique MTurk participants.

We are interested in understanding the norms of information sharing similar to those described by Nissenbaum [109], Criado and Such [28], and Murukannaiah et al. [108] for multiuser scenarios. Therefore, our methodology crowdsources participants' opinions about what information sharing policies are optimal for a number of hypothetical scenarios involving third parties. Shvartzshnaider [128] employ an approach similar to ours, where they crowdsource information sharing norms for an educational context generating hypothetical scenarios involving third parties. We later on show, however, that some scenarios lead to discrepancies in participants opinions and personalisation is the key to accurately recommending sharing policies for multiuser scenarios.

Contributions

1. We propose a novel model for representing and reasoning about multiuser scenarios, employing three types of features: contextual factors, user preferences, and user arguments.

2. We build a classifier, based on machine learning, as an exploratory attempt to predict the optimal sharing policy and evaluate the underlying privacy complexity of multiuser scenarios.

3. Employing the results obtained by the classifier, we identify what combinations of scenario-defining elements introduce variability in judgments on optimal sharing policy.

Organization

Section 6.2 identifies factors potentially influencing a multiuser sharing decision. Section 6.3 describes an inference model for predicting the optimal sharing policy and the MTurk study we conducted to build a dataset for training and testing our inference model. Section 6.4 describes our hypotheses and evaluation strategy. Section 6.5 presents our results and Section 6.6 describes how some of those results can be employed in a practical tool. Section 6.7 discusses our results, threats to their validity, and pointers for future work. Section 6.8 describes related works and Section 6.9 provides conclusions.

6.2 Factors Influencing a Multiuser Sharing Decision

We identify three important classes of factors that potentially influence the privacy decision in a multiuser scenario: context, user preferences, and arguments by users.

6.2.1 Context

Our definition of context reflects intuitions compatible with those of Nissenbaum's [109] study of social settings that dictate the flow of personal information. Her definition of contextual integrity captures the idea that people share information of varying type and sensitivity in society, not simply as individuals in an undifferentiated social world, but as individuals in certain capacities (roles), in distinctive social contexts, such as health care, education, and employment. Since Nissenbaum formulates contextual integrity in the field of law and public policy, legal elements, such as obligation or entitlement, are prominent in the existing account. However, we posit that the legal elements do not adequately cover several aspects of sharing decisions in social networks.

Based on the literature, we focus on three elements of context as factors influencing sharing decisions in multiuser scenarios.

Relationships among the individuals The roles of the individuals involved in a transaction are a defining element of context. Also, the relationship type is crucial when making *individual* decisions about privacy in social media [101, 81]. Specifically, people share information differently with friends, family, and colleagues. We hypothesize that relationship types influence how a person perceives a multiuser scenario, because attributes of a relationship such as intimacy may influence the extent to which the parties involved in the scenario respect each other's opinion. For example, following Wisniewski et al.[159], we imagine that a user's friend would respect the user's preferences.

Sensitivity of the information The sensitivity of the information to be shared is known to influence individual sharing decisions in social media [150, 132]. Besides, the judgment of appropriateness of sharing any information on an SNS is user and culture specific. For example, in some cultures, drinking alcohol is taboo. Thus, a picture showing a person drinking can be inappropriate in one culture, but normal in another. Each individual involved in a scenario has a perception of the sensitivity of the information. This perception may affect how important that person thinks his or her view is on the appropriate sharing decision.

Sentiment associated with the information A user may employ social media for self-presentation and sharing as a means of maintaining that presentation [64, 81]. The sentiment conveyed by the personal information a user discloses on social media influences his or her self-presentation. For example, a user may share a picture of his recently broken leg with friends to receive emotional support. Similarly, a user can publicly post a picture of his graduation to obtain public recognition from everyone and congratulations from friends and family. In these examples, both pictures convey extremely different sentiments, nonetheless, both can be worth sharing, although with different audiences, depending on the intentions of the users. Since one of our goals is to observe the influence of arguments (which may convey the intentions of users) on the final privacy decision, we include sentiment as a factor contributing to context. Further, recent research on social media [130, 165] presents methods to classify pictures based on sentiment, and some studies [138] tackle the influence of the sentiment of the information on the level of its disclosure.

6.2.2 Preferences

The goal of each individual in a multiuser scenario is to persuade the sharer to apply the individual's preferred sharing policy to the information in question. The individuals' (including sharer's) preferences may be either compatible or conflicting. For example, Bob's preference of "I do not want my parents to see this picture" is compatible with Alice's preference of "I want only my friends to see it," as long as Bob's parents are not among Alice's friends. Optimally, the final decision to resolve a conflict should respect the preferences of every individual involved to the best possible extent. If all preferences are compatible with each other, the solution is trivial. However, in case of conflict, an acceptable solution may not be evident.

A sharing policy can imply no sharing, sharing publicly, or anything in between. Further, depending on the number of contacts and their types, sharing policies change from one SNS user to another. Given the large space of possibilities, for the sake of simplicity, we limit the sharing policies in our setting to three levels of disclosure.

1. *Share with all*: Anyone on the SNS can access the information. This sharing preference is at one end of the spectrum of sharing preferences.

2. *Share among themselves*: Only the individuals directly connected with the information can access the information. Since the scenarios presented in our study always include a number of individuals who are willing members of a group picture, the case of sharing among themselves equals the case of no sharing. That is, it does not matter whether the picture is shared among themselves on the SNS, because they shared the moment when the picture was taken, and sharing the moment can be more meaningful than sharing it online. Consequently, this preference is the direct opposite of *share with all*. Thus, we place it at the other end of the spectrum.

3. *Share with common friends*: Only common friends of the individuals involved in the scenario can access the information. As explained before, the space of

possibilities is large and varies from user to user. Thus, there are no predefined privacy preferences that cover every possibility. Nonetheless, we consider that *share with common friends* is a reasonable compromise to present a privacy preference that lies in between the two ends of the spectrum represented by the preceding two preferences.

Two of the above sharing policies (share with all and share among themselves) represent the two ends of the spectrum of sharing possibilities and the other option is in between. We note that the sharing preferences Facebook employs in its basic privacy configuration are *only me*, *friends*, and *public*. The three sharing preferences we consider are the same ones Facebook employs, though adapted to sharing in multiuser scenarios. A benefit of choosing settings similar to Facebook, a popular SNS, is that many participants would be familiar with these privacy settings and would understand their privacy implications.

In addition to the three options above, we considered *share with all friends of all the parties* as a sharing preference to include in our study. First, we note that including an additional preference would increase the total of number of scenarios to investigate. Second, a typical SNS user has a large number of friends (e.g., a recent study of two large samples of Facebook users found the mean values of a user's friends in the two samples to be 155.2 and 182.8 [37]). As a result, sharing with all friends of all three parties involved is likely to include many strangers from each party's perspective. Thus, we concluded that, from a user's perspective, the risk-benefit tradeoff of *share with all friends of all the parties* may be similar to *share with all*. Consequently, we decided not to include sharing with all friends of all the parties as an additional preference.

6.2.3 Arguments

An individual involved in a multiuser scenario may employ arguments to persuade the others that his preferred sharing policy should be used, or at least considered for making the final decision. There are potentially many arguments one can employ to negotiate with or persuade another person. Following Walton et al.[149], we understand arguments as instances of argumentation schemes, each scheme representing a form of inference from premises to a conclusion. Walton et al. show that arguments used in everyday conversation fall into a small number of schemes.

We identify four argumentation schemes that can be effective in deciding an optimal solution for a multiuser scenario: *argument from* (i) *good consequences*, (ii) *bad consequences*, (iii) *an exceptional case*, and (iv) *popular opinion*. It is important to note that we neither claim the foregoing as the only possible argumentation schemes applicable in resolving a multiuser privacy conflict nor do we seek to evaluate if these are the best possible schemes. Our objective is to evaluate if arguments, as instances of argumentation schemes, help in deciding the final policy in multiuser scenarios. Our motivation in adopting argumentation schemes is to restrict the arguments to be of a few well-defined types, instead of choosing the arguments arbitrarily.

The general structure and examples for the argumentation schemes we use is as follows.

Argument from good consequences *If A is brought about, then good consequences will occur. Therefore, A should be brought about.*

An example of an argument from good consequences is: *We had a lot of fun during the party. Everybody's talking about how funny you were and they want to see your pictures. Let's share this picture with everybody.* An SNS user, who shares something, expects to receive some benefit, e.g., friendship, jobs, or other social opportunities [38]. Thus, it is reasonable to argue that sharing certain information implies a good consequence.

Argument from bad consequences *If A is brought about, then bad consequences will occur. Therefore, A should not be brought about.*

An example argument for bad consequences is: *It's a funny picture, but embarrassing since I appear drunk. I don't want strangers seeing it.* Sharing inappropriate information can hurt people's feelings and cause social tension. Thus, negative consequences can be valid arguments for not sharing.

Argument from an exceptional case *If this case is an exception, then the established rule can be waived in this case.*

An example of an argument from an exceptional case is: *C'mon! it was our graduation party! something that we do only once in our lifetimes. We should show it to the world.* Although prior experience can guide future decisions, handling exceptions calls for a different approach. Instances of this scheme cover cases where an unusual privacy configuration may be called for: potentially, the opposite of the policy that might have been adopted if the arguments were not provided. Obviously, an individual must make a strong case to justify that the information is exceptional.

Argument from popular opinion *If a large majority (of the relevant group G) accepts a claim, then there is a presumption in favor of that claim. Similarly, if a large majority rejects a claim, then there is a presumption against that claim.*

We do not consider explicit arguments from popular opinion. That is, in our setting, no user involved in a scenario provides an argument such as: "*the majority of us wants to share this picture only with common friends, hence, we should share it only with common friends.*" Instead, the popular opinion *emerges* when two or more users suggest the same sharing policy. Thus, although no individual explicitly employs an argument from popular opinion, the preferences of the individuals involved in a scenario may imply a popular opinion.

6.3 Inference Model

We envision a computational model that incorporates the factors identified in Section 6.2 to recommend an optimal sharing policy for a multiuser scenario. Figure 6.1 shows an overview of the model.

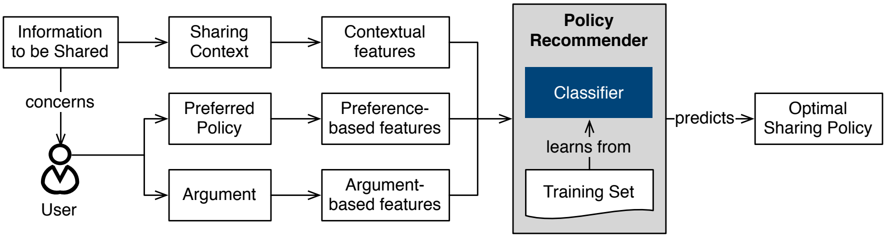


Figure 6.1: An overview of our envisioned inference model that predicts optimal sharing policy in a multiuser scenario

Given a piece of information to be shared in a multiuser scenario, we first identify all users involved in the scenario. This can be accomplished, for example, by enabling a tagging mechanism. We then ask each user identified to provide his or her preferred sharing policy for the information and an argument justifying that preference. As for the sharing context: 1. relationships can be obtained directly from the SNS; 2. users can be asked to rate the sensitivity of the information; and 3. several automated approaches exist for computing the sentiment of the information.

We note that all information above may not be available in a multiuser scenario. For example, some of the individuals involved in a multiuser scenario may not specify their preferences or arguments. Thus, we envision an inference model recommends a sharing policy based on available information and improves the recommendation as more information becomes available.

The core of our model is a *policy recommender*, which employs a machine learned *classifier* to recommend an optimal sharing policy for a given scenario. The recommender can employ a traditional classifier. However, two important challenges are 1. to engineer the features of the classifier based on the sharing context, and the preferences and arguments of the users involved, and 2. to build a dataset to train and test the classifier.

Next, we describe a crowdsourcing approach to build a dataset for training the policy recommender. This approach is useful not only for testing the recommender, but also for building a seed training set required for practical deployment of the recommender. Once the recommender is in use, additional instances can be added to the training set during deployment—as users share information via the model by adopting (or disregarding) the recommender's suggestions.

6.3.1 Data Collection via Crowdsourcing

The dataset we seek must consist of a variety of multiuser scenarios and the optimal sharing policy in each of those scenarios. We expect that, in a multiuser scenario, one party's preferred sharing policy will rarely be the sharing policy all the users involved adopt. Therefore, to collect actual first-party data, we would need to recruit cliques of participants that already know each other and ask them to recreate multiuser scenarios. This is nontrivial and challenging. Users are often reluctant to share sensitive information (one of the contextual factors in our model), biasing the study toward nonsensitive cases that users are willing to reveal [150]. An alternative is asking users to self-report how they behave when they experience a multiuser scenario, but the results may not match participants' actual behavior because of the well-known dichotomy between users' stated privacy attitudes and their actual behavior [1].

Considering the challenges above, we chose to create *situations* in which participants are *immersed* [100] to improve behavior elicitation while avoiding biasing the study to nonsensitive situations. We present information about two or more individuals in a specific circumstance: a combination of context, preferences, and arguments. We ask participants to choose an optimal sharing policy for that circumstance.

An approach similar to ours is to ask participants to imagine themselves as involved in the presented situation. For example, showing a picture of some people and telling the participant that the picture depicts him or her with some friends. We chose against this approach for two reasons. First, participants may feel awkward imagining themselves in the presented situation (e.g., they may feel “that is silly, I would never do that!”), resulting in participants not immersing themselves effectively in the situation. Second, we have multiple users in each scenario and the optimal policy may depend on the particular user’s role we ask the participant to play. This will require us to collect data from the perspective of each user involved in a scenario, increasing the number of scenarios to collect data from three fold.

We recruited participants for our study from Amazon MTurk [110]. We directed each participant to an external website that asked the participant to complete seven survey instruments: a presurvey questionnaire about demographics, five picture surveys (each involving a privacy conflict scenario and three sets of questionnaires), and a post-survey questionnaire about the participant’s general opinions about resolving multiuser privacy conflicts. The study was approved by the IRB at North Carolina State University (details about participants and rewards are in Section 6.3.1.4).

6.3.1.1 Presurvey Questionnaire

We asked participants to report their age, gender, level of education, how frequently they use social media, and how often they share (multiuser) pictures online. Since some of the situations we presented to the participants could be inappropriate for young readers, we required participants to be older than 18 years of age and showed a disclaimer at the beginning that the survey may be inappropriate for some users.

6.3.1.2 Picture Survey

The picture survey is the core of our study. We first show a picture and describe a hypothetical scenario in which the picture was taken and next ask a series of questions. Table 6.1 shows two examples of picture survey. We generated these and several similar picture surveys by combining factors identified in Section 6.2, as described below.

1. Regarding context variables, we consider a predefined set of relation types, namely, *friends*, *family*, and *colleagues*. Squicciarini et al.[134] and Toch et al.[148] employ the same three types as an approximation to a user's relationships on a social network. Further, we assume that all individuals involved in a scenario have the same type of relationship with each other (i.e., all of them are either *friends*, *family*, or *colleagues*). Also, the pictures shown in the situations could be sensitive or nonsensitive and convey either a positive or a negative sentiment. This leads to 12 possible contexts. We found a representative picture for each of those combinations.

   Although we selected these 12 representative pictures, it is important to note that we ask participants to identify the contextual factors for each picture shown to them as indicated in Table 6.1. Specifically, we ask participants to identify the type of relationship among the people involved in the scenario (family, colleagues, or friends), and rate sensitivity (Likert scale 1 = not sensitive at all, 5 = very sensitive), sentiment (1 = extremely positive, 5 = extremely negative). Appendix 6.9 includes all the 12 pictures we employed in our study along with participants' ratings of contextual factors for those pictures.

2. We assume that (1) an *argument from bad consequence* does not often support a *share with all* policy, and (2) an *argument from good consequence* does not often support a *share among themselves* policy. Although such combinations are conceivable, e.g., in scenarios where one of the users involved wishes a negative outcome for another, we posit that such scenarios are not common in real life. We exclude these combinations for simplicity. Further, we restrict the *argument from exceptional case* to support only policies at either extreme: *share with all* or *share among themselves*. Thus, we consider six policy-argument combinations.

3. We limit the number of individuals involved in each scenario to three. This way,

Picture		
Description	Aiko (C) took the picture above with her colleagues Ichiro and Katsu and, a French volunteer at the tsunami relief center	Three friends, Mark, Alex, and John, took the picture above during Mark's bachelor party on a boat in Ibiza
Rating	Identify the relationship between Aiko, Ichiro, and Katsu and rate the sensitivity and sentiment of the picture	Identify the relationship between Mark, Alex, and John and rate the sensitivity and sentiment of the picture
Context	Consider that Aiko wants to upload this picture to her social media account. What sharing policy should she apply for the picture?	Consider that Alex wants to upload this picture to his social media account. What sharing policy should he apply for the picture?
Preferences	Next, consider users' preferences as follows **Aiko** Share among ourselves **Ichiro** Share among ourselves **Katsu** Share with all Considering the context and users' preferences, what sharing policy should Aiko apply for the picture?	Next, consider users' preferences as follows **Mark** Share among ourselves **Alex** Share with common friends **John** Share with all Considering the context and users' preferences, what sharing policy should Alex apply for the picture?
Arguments	Finally, consider the users' preferences and arguments as follows **Aiko** This was one of the worst natural disasters ever. Share among ourselves **Ichiro** Tsunami was a disaster and our hand gestures are not appropriate; people may get the wrong idea. Share among ourselves **Katsu** The picture shows the difficult situation of survivors; sharing this picture can encourage people to help. Share with all Considering the context, and users' preferences and arguments, what sharing policy should Aiko apply for the picture?	Finally, consider the users' preferences and arguments as follows **Mark** There were some girls at the party; people might understand things the wrong way. Share among ourselves **Alex** This was one of the best day of our lives. Share with common friends **John** The is not like any other picture; it was from Mark's bachelor party! Share with all Considering the context, and users' preferences and arguments, what sharing policy should Alex apply for the picture?

Table 6.1: Two example picture surveys (shortened version of those we used)

the implicit *argument from popular opinion* could work (when applied) without ties. Although some scenarios we employed showed pictures with more than three individuals, our scenarios discussed the preferences and arguments of only three of the individuals among the people involved with the picture.

4. We make sure that not all three individuals in a scenario use the same policy-argument combination. Our objective is to understand how a user decides a final policy given the scenario, and the preferences and arguments of others in the scenario. If all users thought the same way and wanted the same result, the solution would be trivial.

Putting the above together, we have: 12 pictures based on context, six policy-argument combinations each of first two individuals can employ, and five policy-argument combinations the last individual can employ (last restriction above). That is, we generated 2,160 scenarios. Each MTurk participant was shown five unique scenarios, making sure that no participant was shown the same picture twice. Scenarios were shown in random order to counter ordering bias. Further, we asked participants to immerse themselves in the particular scenario and ignore the resemblance or lack of resemblance of each scenario to other scenarios in which they might have seen that picture.

Following the picture and its description, we asked participants to identify the contextual factors for the scenario and answer three sets of questionnaires. We asked participants to answer these questionnaires sequentially and when answering a questionnaire, to consider only information provided to them up to that point.

Each of the three questionnaires tells participants that one of the individuals in the scenario wants to upload the picture to a social media account and asks participants what sharing policy should be applied. The participants choose one of the policies from *share with all*, *share with common friends*, and *share among themselves*. In the first questionnaire, participants know only the contextual attributes, but not the preferences or arguments of the individuals in the scenario. This case is similar to a real scenario where a user wants to upload and share information without asking others potentially concerned with the information. The second questionnaire introduces the preferences of all the users, but without their arguments supporting preferences. The third questionnaire employs all of the elements: the individuals in the scenario expose their preferences and support them with arguments. We keep the preferences fixed from the second questionnaire to the third, so we can observe the effect arguments have on the final decision.

6.3.1.3 Post-Survey Questionnaire

The post-survey questionnaire asks the following questions.

1. How important do you think the following factors are in choosing an appropriate policy when sharing information concerning multiple users on social media? (a) Relationship between stakeholders; (b) Sensitivity of the information shared; and (c) Sentiment of the information shared. The response to each factor was on a Likert scale (1 = not important at all, 5 = extremely important).

2. How confident will you be in choosing an appropriate policy for sharing information concerning multiple users on social media in the following cases? (a) You do not know the users' preferences or arguments; (b) You know the users' preferences, but not their arguments; and (c) You know the users' preferences and arguments. The response to each case was on a Likert scale (1 = not confident at all, 5 = extremely confident).

Responses to the above questions enable us to find correlations (or lack thereof) between participants' self-reported behavior and what they actually answered during the study.

6.3.1.4 Participants and Quality Control

We needed 432 participants to receive one response per scenario (each participant responds to five of the 2,160 scenarios described above). We intended to obtain two responses per scenario for completeness. However, we anticipated that some participants would begin a survey but not finish it, leaving gaps in the completed responses. To address this challenge, we launched the study on MTurk in multiple batches. For each batch, we checked if a particular survey needed additional responses and restricted the posted tasks accordingly. The final number of unique participants that completed the study was 988. At the end, each scenario had received at least two responses and some had received three responses. Compensation was provided for only those who completed all seven steps in the survey.

For quality control, we required participants to have completed at least 50 tasks on MTurk and to have had a success rate of at least 90% [114]. We included an attention check question [50] in the ratings section of each picture survey, asking how many people (faces) were present in the picture, answering which requires counting from the picture. Participants answered the attention question incorrectly in a total of 38 instances (less than 1% of responses). If a participant incorrectly answered the attention check question in a picture survey, we excluded that picture survey from analysis, and retained only those picture surveys where the participant answered the attention check questions correctly.

Table 6.2 summarizes our participants' responses to the presurvey questionnaire. The question corresponding to the last row in the table was in the post-survey questionnaire, so that participants understand what multiuser conflicts look like before answering that question. As shown, the majority of our participants used social media on a daily basis. Over 80% of our participants had shared a picture showing multiple users and about one-third of them had experienced privacy conflicts.

Gender	Male: 46.3%, Female: 53.4%, Other: 0.3%
Age	18–20: 2%, 21–29: 36.6%, 30–39: 36%, 40–49: 13.7%, 50–59: 7.5%, 60 or more: 4.1%
Education	Graduate degree: 11.2%, Bachelor degree: 44.4%, College no degree: 30.9%, High school: 12.4%, Less than high school: 1%
Social media usage	Daily: 83.9%, Weekly: 12%, Monthly: 3.7%, Never: 0.4%
Pictures shared	Many (>5): 35.1%, Few (1–5): 45%, None: 18.1%, Not sure: 1.7%
Conflicts experienced	Many (>5): 2.8%, Few (1–5): 30.1%, None: 66%, Not sure: 1.1%

Table 6.2: Demographics of MTurk participants of our study

6.3.2 Building a Training Set

We map the data collected in the MTurk study to a training set the policy recommender learns from. A training set consists of a set of data instances, where each data instance consists of a response variable (class) and a set of predictors (features). We set these as follows.

- A data instance corresponds to a participant's response to a picture survey.
- The response variable is the final policy chosen by the participant.
- The predictors are divided into three cases based on the picture survey. The first case consists of contextual features; the second case consists of contextual and preference-based features; and the third case consists of contextual, preference-based, and argument-based features. For brevity, we refer to these cases as Context, Preferences, and Arguments, respectively.

6.3.2.1 Contextual Features

We compute the contextual features based on participants' responses in the ratings section of the picture survey.

1. *Sensitivity* and *sentiment* each yield a feature with five levels corresponding to integer ratings from 1 to 5.
2. *Relationship* yields a feature with three levels: *family*, *friendship*, and *colleagues*.

6.3.2.2 Preference-Based Features

We compute the following features based on the preferences portrayed in the corresponding scenario (picture survey). For concreteness, we base our examples on Table 6.1 (left).

1. *Preference counts* represents three features corresponding to the number of participants preferring each of the three policies. In Table 6.1 (left), the preference counts are: *share with all* is 1, *share with common friends* is 0, and *share among themselves* is 2.
2. *Most and least restrictive policies* represent, among the preferred policies of the users in a scenario, the policy restricting sharing of information the most and the least, respectively. The order of policies from least to most restrictive is: *share with all*, *share with common friends*, and *share among themselves*. In Table 6.1 (left), the most restrictive policy is *share among themselves*, and least restrictive policy is *share with all*.
3. *Majority policy* represents the policy preferred by the majority of the users involved in the scenario. This feature can be *null* if there is no majority. The majority policy in Table 6.1 (left) is *share among themselves*.

6.3.2.3 Argument-Based Features

We compute argument-based features from the arguments users involved in a multiuser scenario provide for their corresponding preferences (also, recall from Section 6.2.3 that we restrict arguments to be instances of one of the argumentation schemes we consider).

1. *Argument counts* represent the number of times each type of argument is employed in the scenario. For Table 6.1 (left), the counts are: *argument from an exceptional case* supporting *share among themselves* is 1 (Aiko's), *argument from negative consequence* supporting *share among themselves* is 1 (Ichiro's), *argument from positive consequence* supporting *share with all* is 1 (Katsu's), and each of the remaining three argument-policy combinations is 0 (recall from Section 6.3.1.2 that there are six argument-policy combinations).

2. *Arguments supporting least restrictive policy* are the types of arguments that support the least restrictive policy. For Table 6.1 (left), the argument supporting least restrictive policy is *argument from positive consequence* supporting *share with all.*

3. *Arguments supporting most restrictive policy* are the types of arguments that support the most restrictive policy. For Table 6.1 (left), these are *argument from positive consequence* and *argument from an exceptional case* both supporting *share among themselves.*

4. *Arguments supporting majority policy* are the types of arguments that support the policy preferred by the majority of users. For Table 6.1 (left), these are *argument from positive consequence* and *argument from an exceptional case* both supporting *share among themselves.*

When arguments from distinct schemes support a policy, as in the arguments supporting *least restrictive policy* and *majority policy* cases above, we employ the combination of arguments as a distinct feature value. Also, if a majority policy does not exist, we set the corresponding argument-based feature to *null*.

6.4 Evaluation

In this section, we describe our hypotheses and the techniques we use to evaluate those hypotheses.

6.4.1 Hypotheses

1. *H-Influence-Context*: Contextual factors, specifically, sensitivity and sentiment of the information shared, and the relationships among individuals involved influence the choice of optimal sharing policy in a multiuser scenario.

2. *H-Influence-Preferences*: Preferences of users involved in a multiuser scenario influence the choice of optimal policy for the scenario.

3. *H-Influence-Arguments*: Users' arguments for their preferred sharing policies influence the choice of optimal policy in a multiuser scenario.

4. *H-Prediction-Preferences*: Adding preferences to the contextual factors enhances the accuracy of an inference model predicting the optimal policy for a given multiuser scenario.

5. *H-Prediction-Arguments*: Adding arguments to preferences and contextual factors enhances the accuracy of an inference model predicting the optimal policy for a given multiuser scenario.

6. *H-Confidence-Preferences*: Adding preferences to the contextual factors enhances a user's confidence in choosing the optimal policy for a given multiuser scenario.

7. *H-Confidence-Arguments*: Adding arguments to preferences and contextual factors enhances a user's confidence in choosing the optimal policy for a given multiuser scenario.

6.4.2 Evaluation Strategy

6.4.2.1 H-Influence-$*$

To evaluate our hypotheses about influences of different factors, we build *multinomial logistic regression* models (multiple predictors and one response variable). We adopt Akaike Information Criterion (AIC) as a measure of *goodness of fit* for these models.

$$\text{AIC} = 2k - 2\ln L, \tag{6.1}$$

where L is the maximum value of the likelihood function for the model, and k the number of estimated parameters in the model. Lower values of AIC indicate better fit. AIC rewards goodness of fit (as assessed by the likelihood function), but includes a penalty that is an increasing function of the number of estimated parameters. The penalty discourages overfitting—increasing the number of parameters in the model almost always improves the goodness of the fit [55].

Further, we focus on *coefficients* and their statistical significance. These coefficients help us understand each feature's relative influence on the optimal policy. We employ *share among themselves* as the reference category for all models.

Linear regression models use the general linear equation $Y = b_0 + \sum b_i X_i$, where Y is a continuous response variable and b_i is the coefficient of predictor X_i. However, in logistic regression models, the coefficients are expressed in log-odds units. The coefficients show the effect of a predictor on the log odds of the optimal policy being in a given category ($Y = 1$) versus being in the reference category ($Y = 0$); that is, the odds of Y being a given category increase by a factor of e^{b_i} per unit change in X_i. The equation used by the logistic models is

$$P(Y=1) = \frac{1}{1+e^{-b_0-\sum b_i X_i}} \tag{6.2}$$

We employ dummy variables for categorical predictors (e.g., relationship types). In these cases, the reference category is the one that is missing in the table reporting the coefficients of each model.

Since we consider several hypotheses in each model, the probability of a Type I error (false positive) is high. To reduce these errors, we apply the Holm-Bonferroni correction [67] during the evaluation of the statistical significance of the coefficients yielded by the models.

We include all the features of a type (contextual factors, preferences, and arguments), when possible. However, some features of the same type can be highly correlated, causing *multicollinearity* [59]. In that case, the coefficients may change erratically in response to small changes in the data. To counter this effect, we create independent models for highly correlated predictors. Note that this correction usually leads to higher coefficients.

Further, since each participant responded to a subset of scenarios (five out of 2,160), the personal privacy attitudes of participants may affect the estimated coefficients of the models. Thus, we employ mixed modeling [121] to create the logistic regression models. Mixed models offer the possibility of grouping samples of the data hierarchically. Specifically, we group responses by participants by employing the participant ID as a *random effect*. In this way, the participant ID captures the variability introduced by personal privacy attitudes in the responses. Introducing the ID enables us to obtain estimated coefficients for the other predictors (*fixed effects*) in a way that is less affected by variability in the participants' responses.

6.4.2.2 H-Prediction-$*$

To evaluate our hypotheses about prediction, we build a *classification* model. We evaluate the prediction accuracy of these models via:

$$\begin{aligned} \text{precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ F_1\text{-measure} &= 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}, \end{aligned} \tag{6.3}$$

where TP, TN, FP, and FN refer to true and false positives and negatives. We implemented these models using Weka [60]. We perform ten-fold cross-validation and cross-validated paired t-test [33] to test significance in differences, when required.

We note that the data we collect contains sharing policies that are considered optimal from the view points of third parties. Participants' responses may have been different if they were personally involved in the scenarios shown during the study. Nonetheless, the data and results obtained from the study are still useful to investigate what elements influence the optimal sharing policy and to build an initial dataset to train the policy recommender shown in Figure 6.1. However, in practice, to achieve a high prediction accuracy, this recommender would require adaptive capabilities to learn user-specific privacy behavior and predict sharing policies accordingly.

6.4.2.3 H-Confidence-$*$

To evaluate our hypotheses about users' confidence, we perform the Kruskal-Wallis test [66] on self-reported data from the post-survey questionnaire. The Kruskal-Wallis test is a nonparametric extension of one-way ANOVA and it relaxes the assumption that the populations underlying the samples compared are normally

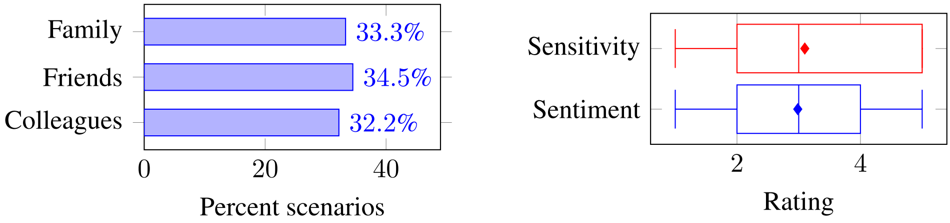


Figure 6.2: Distributions of context variables computed from MTurk participants' responses

distributed. This test compares medians to determine if all ratings come from the same distribution. If the Kruskal-Wallis test determines that all the ratings do not come from the same distribution, we perform the multiple comparisons test [63] to determine which variables are significantly different. As before, since we test multiple hypotheses, we employ the Holm-Bonferroni correction to reduce Type I errors.

6.5 Results

In the following sections, we evaluate each of our hypotheses.

6.5.1 Context (*H-Influence-Context*)

Recall that we ask participants to identify the contextual factors for each picture shown to them. Contextual factors based on participants' responses are what we employ in all of our analyses. Figure 6.2 shows the distributions of the contextual variables computed from participants' responses. These distributions suggest that our picture surveys represent quite some variety in terms of relationship, sensitivity, and sentiment.

Table 6.3 shows the coefficients of the multinomial logistic regression models employing only contextual factors as predictors. The table shows the coefficients we obtain for models of optimal policies in three cases. Recall that in each case the participants provided the optimal sharing policy by considering a different set of factors: Case 1, considering only the contextual factors; Case 2, considering contextual factors and preferences; and Case 3, considering contextual factors, preferences, and arguments. The coefficients for each case are shown in two columns: one for the coefficients of the optimal policy being *share with all* versus the reference category *share among themselves*, and the other for the coefficients of the optimal policy being *share with common friends* versus the reference category. We highlight statistically significant differences at significance levels (α) of 5% and 1% with * and **, respectively. As mentioned before, where multiple hypotheses are tested, we adjust the significance levels via the Holm-Bonferroni correction.

In models for all three cases, we find that sensitivity is a significantly influential factor. Specifically, in the first case, where only contextual factors are considered, the estimated coefficient -5.485 indicates that the probability of the optimal sharing policy being *share with all* compared to the probability of being *share among themselves* decreases $e^{-5.485}$ times for each increase in the sensitivity level of the picture, assuming everything else remains unchanged.

The results also indicate that the significance of the type of the relationship decreases noticeably when preferences and arguments are considered (Cases 2 and 3). For example, in both Cases 2 and 3, *relationship = colleagues* is not statistically significant on the relative probability of the optimal policy being *share with all* versus *share among themselves*. Similarly, *relationship = friends* is not significant for the optimal policy being *share with common friends* versus the reference category. In addition, sentiment is not significant in any case.

Our intuition, based on these observations, is that, most of the times, users consider it appropriate to share a highly sensitive picture only among the individuals involved in that picture.

Further, we observe that as participants consider preferences and arguments, the coefficients of the contextual factors decrease and AIC increases. This indicates that, when users take preferences and arguments into account, the contextual factors'

Variable	Coefficients					
	Context (Case 1)		Preferences (Case 2)		Arguments (Case 3)	
	All	Common	All	Common	All	Common
Sensitivity	−5.485**	−4.174**	−2.045**	−1.508**	−1.707**	−1.38**
Sentiment	−0.415	−0.424	−0.352	0.057	−0.193	0.048
Rel = Colleagues	−0.525**	−1.345**	−0.355	−0.457**	−0.245	−0.543**
Rel = Friends	1.162**	−0.464**	0.507**	−0.165	0.659**	−0.154
AIC	2481.1	3411.5	2471.0	4456.0	2599.0	4425.2

Table 6.3: Regression coefficients for contextual variables

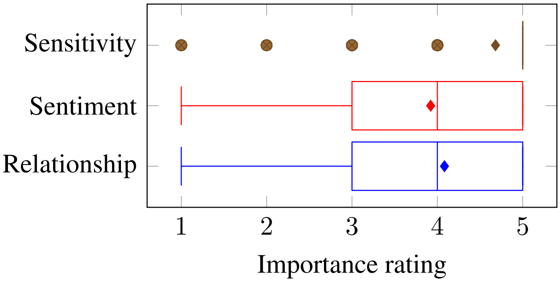


Figure 6.3: Importance (from the post-survey questionnaire) of context variables

influence on the optimal sharing policy diminishes.

To further analyze the contextual factors, we compare these results with self-reported data collected in the post-survey questionnaire. In this way, we can observe if participants' opinions are consistent with the decisions they made in the picture survey. Figure 6.3 shows the boxplots for participants' importance ratings to each contextual factor. A diamond dot in the boxplot indicates the mean. Each dot outside a box indicates an outlier.

From Table 6.3 and Figure 6.3, we observe that the regression model and self-reported values follow a similar pattern: sensitivity is the most influential factor. However, an important distinction between the two is that although relationship and sentiment have similar ratings in the post-survey questionnaire, sentiment is not significantly influential in the regression model. Also, in the post-survey questionnaire relationship and sentiment have high (median of 4) importance ratings.

Variable	Coefficients			
	Preferences		Arguments	
	All	Common	All	Common
# Share with all	1.32**	0.693**	1.374**	0.767**
# Share with common friends	0.136	1.029**	0.119	0.978**
# Share among themselves	−1.296**	−1.468**	−1.322**	−1.494**
Least restrictive policy = All	1.021**	0.013	1.197**	0.229
Most restrictive policy = Self	−1.421**	−2.461**	−1.31**	−2.304**
Majority policy = All	0.674**	−0.452*	0.766**	−0.538**
Majority policy = Common	−0.174	−0.195	−0.054	−0.253
Majority policy = Self	−0.584*	−1.07**	−0.536	−1.173**
AIC	2309.0	3849.2	2380.2	3820.1

Table 6.4: Regression coefficients for preference features

However, their low regression coefficients suggest that participants did not consider relationship and sentiment quite as important as they reported in the post-survey questionnaire.

6.5.2 Preferences (*H-Influence-Preferences*)

The logistic models for the contextual factors suggest that employing preferences can influence the final sharing decision. Table 6.4 shows the coefficients for models employing preference-based features. It shows only Case 2 (Preferences) and Case 3 (Arguments) since Case 1 (Context) does not include preferences. We note that preference counts are highly correlated: when one count increases, the other two counts (naturally) decrease, causing multicollinearity. To counter this effect, we create independent models for the highly correlated predictors. Thus, the coefficient values for the preference counts features are obtained from different regression models: each of these models employs only one preference count as its predictor.

First, we observe that the most restrictive policy has the most influence on the final policy, more so than even the majority policy. Second, we observe that preference counts have, in almost every case, a statistically significant influence on the final sharing policy. Third, the coefficients of the preference-based features do not change much from Preferences to Arguments. This lack of change indicates that preferences remain important even when both preferences and arguments are considered.

6.5.3 Arguments (*H-Influence-Arguments*)

Table 6.5 shows the regression coefficients for the argument-based features. Just as the policy counts are highly correlated, so are the argument counts. Thus, we use different models for those features.

Considering the absolute values of the coefficients (ignoring the sign), preferences supported by exceptional and positive arguments have the highest influence in the optimal sharing policy. In contrast, the arguments for negative consequences yield the lowest coefficients, suggesting that our participants weighed the benefits of sharing more than the potential risks of sharing a picture.

However, it is worth noting that *argument from bad consequences* supporting *self* (share among themselves) has the highest coefficient of all argument counts. This observation indicates that when a user makes a strong case for not sharing a picture, the other users respect that preference. This finding is consistent with those of Besmer and Lopford [6] and Wisniewski et al.[159].

6.5.4 Prediction (*H-Prediction-Preferences* and *H-Prediction-Arguments*)

The foregoing hypotheses concerned how various features influence a sharing policy. Now, we evaluate a predictive model that puts these features to use. Since our objective is to predict the actual policy (*all*, *common*, or *self*), we build a three-class classifier, which makes discrete predictions.

Figure 6.4 shows the distribution of sharing decisions opted by our participants in the

Variable	Coefficients	
	All	Common
# Positive supporting all	1.883**	1.071**
# Positive supporting common	0.534**	1.475**
# Negative supporting common	−0.214	1.149**
# Negative supporting self	−2.713**	−3.436**
# Exceptional supporting all	1.274**	0.727**
# Exceptional supporting self	−1.066**	−1.197**
Positive supporting least restrictive policy	−1.612**	−1.257**
Negative supporting least restrictive policy	−1.728**	−0.168**
Exceptional supporting least restrictive policy	−1.334**	−1.043**
Positive supporting most restrictive policy	1.806**	1.354**
Negative supporting most restrictive policy	1.082*	1.269**
Exceptional supporting most restrictive policy	1.532**	1.446**
Positive supporting majority policy	2.641**	1.728**
Negative supporting majority policy	−0.111	0.671**
Exceptional supporting majority policy	−2.385**	−2.53**
AIC	2417.0	3975.6

Table 6.5: Regression coefficients for argument features

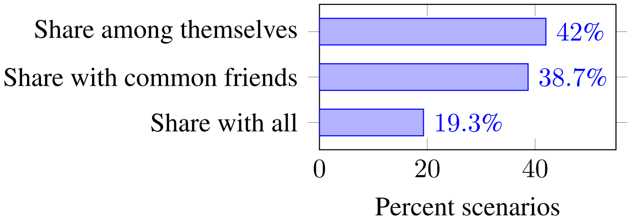


Figure 6.4: Distributions of sharing decisions computed from MTurk participants' responses

	Precision	Recall	F_1	Class	Size
Context	0.539	0.379	0.445	*all*	26.5%
	0.493	0.472	0.482	*common*	36.0%
	0.588	0.736	0.654	*self*	37.5%
	0.541	**0.546**	**0.537**	weighted mean	–
Preferences	0.329	0.047	0.082	*all*	15.4%
	0.606	0.622	0.614	*common*	40.8%
	0.610	0.779	0.684	*self*	43.8%
	0.565	**0.602**	**0.563**	weighted mean	–
Arguments	0.347	0.122	0.180	*all*	16.1%
	0.610	0.640	0.624	*common*	39.2%
	0.646	0.769	0.702	*self*	44.7%
	0.584	**0.614**	**0.588**	weighted mean	–

Table 6.6: Precision, recall, and F_1 scores of logistic regression classifiers for the three cases, considering all data instances

entire dataset. Although participants' preferences were often conservative, we note that participants opted each preference on multiple occasions. This indicates that all three options we provided were useful.

We build *logistic regression* classifiers for different feature sets using the Weka implementation [60]. Table 6.6 compares the precision, recall, and F_1 scores of the classifiers for the three cases. The size column in the table shows, for each class, the percentage of the actual (ground truth) instances belonging to that class.

Analyzing the results obtained by the classifier, we observe that the F_1 measure increases (1) from Context to Preferences: contextual features to contextual and preference-based feature (*H-Prediction-Preferences*), and (2) from Preferences to Arguments: contextual and preference-based features to contextual, preference-based and argument-based features (*H-Prediction-Arguments*). Further, from a 10-fold cross-validated paired t-test [33], we find that differences in the F_1 measures are significant ($p < 0.01$).

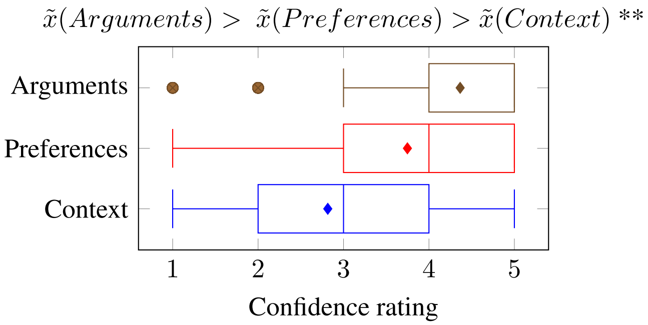


Figure 6.5: Users' confidence (from the post-survey questionnaire) in choosing the optimal policy for different cases

6.5.5 Confidence (*H-Confidence-Preferences* and *H-Confidence-Arguments*)

We compare a user's confidence in choosing an optimal sharing policy given information corresponding to the three cases. Figure 6.5 shows the boxplots of the participants' confidence ratings collected from the post-survey questionnaire. Further, from the Kruskal-Wallis and multiple comparison tests, we find that the median ($\tilde{x}$) confidence rating for Preferences compared to Context, and Arguments compared to Preferences is significantly ($p < 0.01$) larger. This suggests that augmenting the contextual factors with preferences and arguments increases a user's confidence in choosing an optimal sharing policy for a scenario.

6.5.6 Analysis of Misclassified Instances

We investigate the data instances our prediction model (Section 6.5.4) misclassified to find out whether some combinations of scenario-defining elements make the prediction of optimal sharing policy difficult. Because manually identifying such combinations is nontrivial, since our prediction model employs multiple features, we employ two automated techniques for this purpose.

First, we cluster the misclassified instances via the expectation-maximization (EM) algorithm [32]. Our intuition in clustering is that there can be multiple types of scenarios that make prediction difficult. The EM algorithm yields six clusters. We observe that the distribution of scenarios with different ground truth policies (*all*, *common*, and *self*) is almost uniform across the six clusters—none of the six cluster contains more than a 40% scenarios with a given optimal sharing policy as ground truth. Thus, the scenarios in each cluster are not linked to a specific optimal sharing policy (and so are our following observations about the clusters).

Next, to report the clusters visually, in a human-readable format, we build a decision tree with the six clusters as decision targets. The decision tree yields nearly perfect accuracy (99.93%). Figure 6.6 shows the resulting decision tree, where each leaf represents a cluster and shows its name and size. For brevity, we employ the following notation for representing the nonleaf nodes in the tree:

- *#share*[*All*|*Common*|*Self*] represents the number of users expressing that specific sharing preference in a scenario. For instance, in a scenario, if two users prefer the picture to be shared with common friends, *#shareCommon* equals 2.
- *#arg*[*Positive*|*Negative*|*Exceptional*], represents the number of users employing a given argument to support their preferences.
- [*most*|*least*]*RestrictivePolicy* represents the most or least restrictive policy employed in a given scenario. For example, in a scenario, if *#shareAll* = 2 and *#shareCommon* = 1, then *mostRestrictivePolicy* = *common*.
- *argTypeFor*[*MostRestrictive*|*LeastRestrictive*|*Majority*]*Policy* represent the types of argument employed in a given scenario to support the most restrictive, least restrictive, or majority policy, respectively.
- *sensitivity*, *sentiment*, and *relationship* represent the values of these contextual factors employed in a given scenario.

Our objective in training the decision tree is to interpret the paths in the tree. Our intuition is that the path from the root to a cluster-leaf in the tree describes the

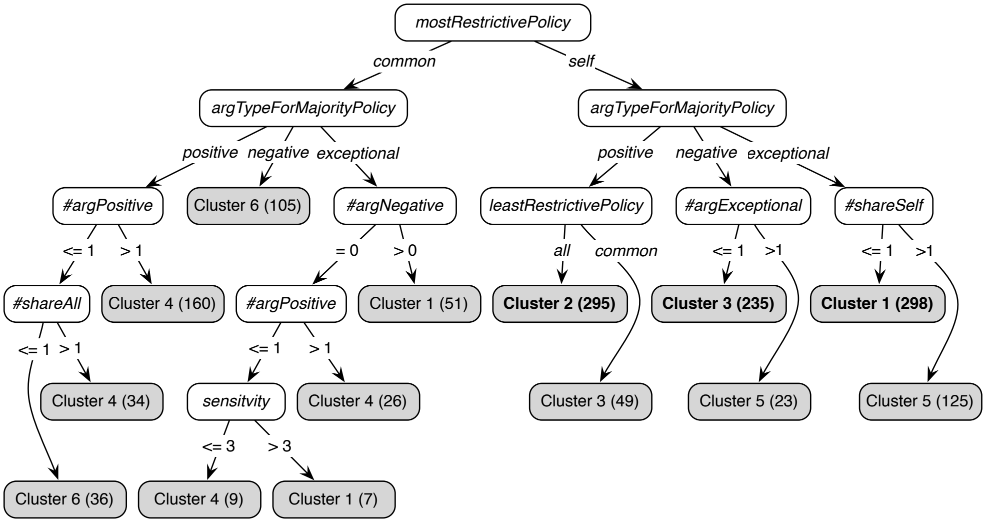


Figure 6.6: A decision tree with the six clusters of misclassified scenarios as decision targets (leaf nodes; shaded).

combinations of feature values that lead to misclassifications corresponding to the cluster.

On the one hand, we observe that few nodes in the tree represent contextual factors. Specifically, there are only two paths that contain *sensitivity*; further, the clusters corresponding to these two paths represent only 16 scenarios. Their rarity suggests that combinations of contextual factors defining a scenario have little effect on whether the scenario is misclassified or not.

On the other hand, we observe that most nodes in the tree (including nodes at the top two levels) are related to users' preferences and arguments in a scenario. This suggests that preferences and arguments can help explain several misclassified instances. By observing multiple paths in the tree, we infer that misclassified scenarios often present some conflict in terms of preferences and arguments. To understand this better, consider three leaf nodes representing Clusters 1, 2, and 3 (highlighted as bold in Figure 6.6), which account for the majority of misclassified instances (61.4%).

- In Cluster 1, the most restrictive policy is *self*, and the majority policy is either *all* or *common* (since #*shareSelf* ≤ 1).
- In Cluster 2, the most restrictive policy is *self* and the least restrictive policy is *all*.
- In Cluster 3, a *negative consequence* argument supports the majority policy, whereas an *exceptional case* argument seems to support the minority policy.

In each of these clusters, a participant must resolve the conflict somehow to choose a final policy. We conjecture that a participant's choice of final policy for a scenario posing conflicts depends on how the participant interprets and more importantly, prioritizes the preferences and arguments involved. Further, since such prioritization is likely to be subjective, two participants may choose different policies as optimal for the same scenario, which can eventually lead to misclassification.

6.6 Muppet: A Tool for Multiuser Privacy Decision Assistance

Despite achieving only moderate accuracy, our prediction model can be a valuable tool for deciding sharing policies in multiuser scenarios. To illustrate this potential benefit, consider how one would employ our model in practice. Imagine a tool, Muppet, for multiuser privacy decision assistance.

Step 1 Given a piece of information to be shared, the Muppet tool identifies users concerned with the sharing of the information. In a simple use case, one of the users involved can manually tag the information with identifiers for all the concerned users. However, this step (or parts of it) can potentially be automated. For example, many picture editors automatically detect and tag faces in a picture; if the information to be shared is a social media post, concerned users are often mentioned (e.g., by including "@Alice") within the post.

Step 2 Muppet identifies the context of the information. Again, one of the users involved with the information can manually identify the context of the information. However, some contextual information can be automatically identified. Specifically, (1) relationship information can be obtained from an SNS provided users organize their connections on the SNS based on relationships, and (2) several automated approaches exist for computing the sentiment for both text and pictures, e.g., [165].

Step 3 Each relevant user specifies their preferred sharing policy for the information and provides an argument supporting their preference. In our current model, a user must choose one of the predefined sharing policies and argument types. For example, a user may choose the preferred policy as *all* and the argument as *exceptional case*.

Step 4 Given the information above (context of the information, users' preferences, and argument), the Muppet tool recommends a sharing policy for the scenario. The concerned user may adopt or disregard the recommendation. However, if the user's final decision can be observed, Muppet model can be retrained in an online fashion, e.g., as in Murukannaiah and Singh [106].

Considering the difficulty of collecting data from first parties, an initial training of our model can be performed on a dataset synthesized from third parties choosing optimal sharing policies in a variety of scenarios (such as the dataset we built from our MTurk study). A downside of this approach is that such a dataset may not accurately reflect users' sharing behavior in multiuser scenarios (since third parties do not share information for real).

An advantage of training with a dataset of third-party views, though, is that the resulting model would be functional off-the-shelf. As explained above, the privacy settings selected by the participants represent socially accepted information-flow norms. This is strengthened by the fact that there was a consensus between participants on the optimal policy for a scenario in the majority of scenarios employed in our study (specifically, for 63% of the scenarios we employed, all participants that rated a given scenario provided the same optimal sharing policy). Therefore, our model can recommend sharing policies starting from the first piece of information a user wants to share via a tool based on our model. Although the model may yield moderately accurate recommendations in the beginning, we posit that its accuracy would increase if the training set is updated with first-party data collected from groups of users, and the recommender retrained accordingly.

Personalizing Muppet

As shown in Table 6.6, a machine learned classifier that employs contextual features, preferences, and arguments achieves a moderate accuracy. However, this classifier ignores the characteristics of the users involved in a scenario. Since sharing norms may vary depending on individuals and groups, we envision personalizing Muppet's recommendations depending on characteristics of the specific users involved in a multiuser scenario. We posit that such a recommender, which learns from and adapts to users, would yield a higher accuracy than the classifier we presented in this paper.

As an initial investigation to understand the influence of personal features on the optimal sharing policy in a multiuser scenario, we build a multinomial regression model employing demographic variables (collected in the presurvey) as predictors and the optimal sharing preference provided by the participants as the response variable.

As shown in Table 6.7, the coefficients for most demographic variables are not statistically significant. However, two main exceptions are gender and sharing experience. Specifically, women are much more reluctant than men to share information openly once they consider the preferences and arguments of the parties involved in the multiuser scenario. This indicates that, in many cases, women dismiss the least restrictive policy proposed in the scenario. Next, the greater the sharing experience a user has, the more inclined that user is to share openly. Based on these

Variable	Coefficients					
	Context (Case 1)		Preferences (Case 2)		Arguments (Case 3)	
	All	Common	All	Common	All	Common
Gender = male	0.486**	0.064**	0.515**	0.121**	0.463**	0.137**
Gender = female	0.541	0.036	−11.74**	−0.463	−11.53**	0.215
Age	−0.101*	−0.067	−0.136**	−0.026	−0.121**	0.003
Education level	−0.051	−0.042	0.057	−0.009	0.068	−0.033
Socialmedia use frequency	0.142*	0.061	0.181*	0.055	0.195*	0.027
Sharing frequency	−0.285*	−0.121	−0.309*	−0.103	−0.324*	−0.061
Sharing experience	0.281**	0.246**	0.294**	0.288**	0.296**	0.273**
Conflict experience	−0.068	−0.102	−0.029	−0.172	−0.031	−0.118
AIC	6127.15	9087.8	5872.8	8493.28	5988.5	8549.93

Table 6.7: Regression coefficients for demographic variables

results, Muppet can include heuristics to personalize recommendations. For example, Muppet may recommend a more conservative policy for a scenario which involves a majority of women compared to a similar scenario with a majority of men.

6.7 Discussion

Our results show that contextual factors, preferences, and arguments influence the optimal sharing policy. However, we find a disparity between the importance ratings participants assigned to contextual factors and how these factors actually influence the optimal policy. Participants considered that the sentiment a picture conveys is an important element (rating 4 out of 5). However, according to our regression model, the influence of sentiment on the optimal policy is not statistically significant.

Our regression and inference models show that arguments help in predicting the optimal policy. Further, the majority of participants reported that they feel much more confident about what policy to apply to a shared item if they know others' arguments. These results indicate that approaches based only on aggregating user preferences, e.g., [147, 20, 70] (as discussed below), may not be effective. Our results show that users are willing to accommodate others' preferences if they know their reasons. Therefore, knowing the preferences alone is not sufficient to set a policy that is satisfactory for the majority of the users involved.

The inference model achieves the F_1 score of 58.6% when trained with features based on contextual factors, preferences, and arguments. Although the accuracy of our model is not high in absolute terms, it performs noticeably better than what some naive inference models can achieve. Specifically, (1) a *random classifier* that randomly chooses one of three sharing policies as the optimal policy would yield a 33.3% accuracy, and (2) a *majority-class classifier* that always chooses the policy corresponding to the majority-class as the optimal policy would yield a 44.6% accuracy (Table 6.6). We are not aware of other models (that predict sharing policies in multiuser scenarios) with which we can compare our model. Our model can serve as a baseline for future works on this topic.

Our analysis of the misclassified instances reflects the complexity of considerations that go into argument evaluation. Our results suggest that arguments are crucial in helping participants conceive the information-flow norms for a given context. However, as our analysis suggests, these norms may be subjective. That is, although some users may consider an information flow in a specific context as a norm, others may consider the same flow in the same context as a privacy violation.

6.7.1 Threats to Validity

We identify two threats to validity of our findings. First, any study based on surveys is susceptible to participant confusion. We mitigated this threat via explicit quality control measures commonly used in MTurk studies, such as selecting workers based on number of previous tasks completed and task success rate, and employing attention-check questions, as detailed in Section 6.3.1.4.

Second, to avoid the unreliability of self-reported attitudes, the well-known reluctance of participants to provide sensitive information, and the logistical challenges of finding participants in sets of three friends, we employ scenarios in which users immerse themselves and provide their assessment of the policies, preferences, and arguments under consideration. This is based on methodologies known to work well in other domains, such as [100], as well as methodologies used in other social media privacy studies, such as [61, 83, 134, 156]. Generalizations, however, should be made with caution, as participant's decisions in such immersive scenarios may not necessarily match those in real scenarios.

6.7.2 Limitations and Directions

We identify a number of directions for future work. For logistical reasons, we employ predefined types of arguments, preferences, and contexts. Thus, our findings are specific to these values. Nonetheless, other relationship types, sharing preferences, and argumentation schemes can be appropriate or even desirable in a multiuser scenario. A future study could attempt to understand what arguments users would employ to support their preferences during a multiuser scenario—the prospective benefit being to discover argumentation schemes geared to multiuser sharing. Automated techniques for identifying arguments [96] can be valuable in this regard.

Similarly, the influence of other elements suggested in the literature on the optimal policy is worthy of investigation. For example, Nissenbaum [109] identifies entitlement as a context-defining factor. In SNSs, users may consider that the owner of the information item is more entitled than anyone else to define the final sharing policy for that item.

We restrict the number of sharing policies to three options (share with all, common, themselves) so as to limit the possible scenarios to a manageable number in our user study. Therefore, another path for future research is to investigate how users solve privacy conflicts when fine-grained sharing preferences are used. Additionally, researchers should look into whether modifying the potential viewers can help users reach a satisfactory agreement. For example, a user who cannot be persuaded through arguments to share a picture with common friends may waive her objection if the other parties agree to exclude a specific common friend from the target audience.

In our study design, to limit the complexity of the multiuser scenarios, we restricted the negotiation among the parties to one round. Wisniewski et al.[159] and Besmer and Lipford [6] identify several social aspects that people employ to figure out an optimal sharing policy (e.g., trust and group norms). A future direction is to incorporate such social aspects in the model presented in this paper to determine how they evolve during successive rounds of a negotiation. For example, arguments can change from round to round depending upon the preferences expressed by others. And, group norms can dictate what arguments can be employed in a negotiation.

Our work contributes by informing the development of decision-support systems for multiuser scenarios by modeling context, preferences and arguments to find an optimal sharing policy. However, additional challenges need to be addressed in order to develop a usable user interface for any such decision-support system, such as finding an adequate trade-off between user intervention and automation to avoid burdening users. To this aim, suitable defaults, e.g., [153], or recommender systems based on machine-learning approaches, e.g., [39], could be applied. As future work, we plan on building a recommender tool and evaluating it with users, so as to collect information about its accuracy and usability.

A user may employ a multiuser decision-support system such as Muppet frequently. For example, sharing group photos may be a daily routine for a user. Therefore, automating the process to a good extent is an important consideration. To this end, we alluded to automated techniques for recognizing sensitivity (e.g., [113]), sentiment (e.g., [165]), and relationships (e.g., [105, 41]). However, this is still an active research area. Thus, a key challenge for future work on multiuser decision-support systems is evaluating not only the recommendations of such such systems but also the automated techniques they build on.

In training and evaluating our policy recommender, we consider overshares (e.g, recommending *share with all* when the optimal is *share with common*) and undershares (e.g., recommending *share among themselves* when the optimal is *share with common*) as equally bad. However, in practice, recommending a conservative sharing policy may be less harmful than recommending a liberal policy when there is doubt on the optimal. Thus, a direction for future work is to consider cost-sensitive models that would weigh overshares and undershares differently.

The findings presented in this paper can be employed in a number of ways. The logistic regression models show how elements of context, preferences, and arguments influence information-flow norms that dictate sharing decisions in multiuser scenarios. Therefore, formal models of norms, e.g., [4], [131], and [29], can be adapted for multiuser scenarios by including those elements. Also, the reported accuracy of our classifier can serve as a benchmark for future classifiers.

Our study is cross-sectional (one time) and in the scenarios of the study, decisions are to be made in one round. Thus, our study does not provide data to understand how users deal with situations where their arguments are countered by others' arguments.

Finally, each participant in our study responded to five scenarios. Although we employed random effects in building the logistic regression models, personal privacy attitudes may bias the results. A future study could require participants to engage in negotiation and persuasion to decide an optimal sharing policy for a scenario. Such a study requires coordinating multiple participants; thus, conducting such a study on MTurk is nontrivial. Employing automated agents (built based on the data we collected in our study) against one or a few real participants is a viable and interesting direction.

6.8 Related Work

As Wisniewski et al.[159] identify, the lack of tools to manage multiuser scenarios forces SNS users to employ a number of *coping strategies*, which are of limited effectiveness in practice. Such strategies include blocking other users, filtering friendship requests by the number of common friends, and self censorship. In a similar previous study, Lampinen et al.[87] explore SNS-users' perceptions of control over online disclosure through a series of interviews. The qualitative data obtained in their study suggests that users manage interpersonal boundaries both individually and collaboratively. Lampinen et al. further classify the strategies that SNS users employ to manage their privacy into two super classes: preventive and corrective. Strategies in both classes can be accomplished through individual or collaborative means. For example, a collaborative preventive strategy is asking for approval before disclosing content from those involved. These two studies show that SNS users need functionalities to manage multiuser privacy. Our work is a stepping stone toward offering such functionality.

Besmer and Lipford [6] propose a method where the owner of a picture in a multiuser scenario is in charge of deciding the sharing policy and the other users involved can suggest privacy preferences. They developed a Facebook application that enables a user to send privacy suggestions to the owner of a picture where that user appears. Besmer and Lipford show that the owners of pictures usually entertain and accept suggestions from others when they decide sharing policies—which is consistent with our findings. A shortcoming of their approach is that it is *manual*, i.e., the owner of a picture must determine an optimal solution based on suggestions from others. Such manual effort may overload the owner of the picture.

Thomas et al.[147] propose *veto voting* as a direct approach to manage multiuser sharing policies. That is, denying access takes precedence over granting access. Thus, if an individual wants to share some information with a given user, but another individual does not, the information is not shared. Whereas this approach protects privacy, it may lead to utility loss. For example, suppose Alice and Bob appear together in a picture. Bob initially opposes sharing the picture with Charlie as he does not know him. However, if Alice tells him that Charlie is her friend and that everything is OK, then Bob may accept sharing with Charlie. Had veto voting been applied, the picture would have not been shared with Charlie, thereby missing, for example, a potential opportunity for Bob to be friends with Charlie.

Other approaches too tackle multiuser privacy conflicts. However, they exhibit various limitations, which we describe next. Some of these approaches require too much human intervention during conflict resolution: Wishart et al.[158] require users to solve the conflicts *manually*. Likewise, Squicciarini et al.[133] requires users to resolve conflicts nearly *manually* by participating in difficult-to-comprehend auctions with fake money to handle every possible conflict.

Approaches that provide automated support [20, 70, 147] help resolve multiuser conflicts, but they usually consider only one fixed way of aggregating user preferences, without considering how users would compromise and the concessions they might be willing to make in a specific situation. Hue et al.[71] consider more than one way of aggregating user preferences, but the user who uploads an item chooses the aggregation method to be applied, which becomes a unilateral decision without considering any input from others. Clearly, solutions that do not consider input from the other users involved may lead to solutions that are far from what some users would be willing to accept. This may lead users to manually resolve conflicts most of the time. Such and Criado [142, 143] provide an improvement over the fixed ways of aggregating user preferences by automatically inferring the particular situation for the conflict and applying the *concessions* that are known to happen during offline negotiations in those situations [87, 159]. Situations are modelled considering the individual preferences of each user involved, the sensitivity of the content, and the relationships to the potential audience. While this approach captures and adapts to known situations, it may not capture opportunistic concessions or agreements that may arise in potentially unknown situations.

Some recent works propose game-theoretic mechanisms to tackle multiuser privacy conflicts. They define negotiation protocols, which are a means of standardising the communication between participants in the process of negotiating a solution to a multiuser privacy conflict by defining how the participants can interact with each other. These protocols are then enacted by users themselves manually [72] or automatically by software agents [144] to negotiate an agreed sharing decision for a particular item. Participants can follow different strategies when enacting the negotiation protocols, and these strategies are analysed using well-known game-theoretic solution concepts such as the Nash equilibrium. However, such proposals may not work well in practice since they assume users are perfectly rational and do not capture the social idiosyncrasies that users consider in real life [87, 159].

Ilia et al.[73] present a mechanism to enforce fine-grained access control in pictures by blurring the faces of the users depicted in the picture based on each users' access control list. This approach can limit the utility of sharing information. However, used in conjunction with a decision-support mechanism based our findings, Ilia et al.'s approach could enforce access control differently for different users in case they cannot agree on a sharing policy.

Arguments are commonly used to resolve conflicts in other domains. Murukannaiah et al.[107] employ arguments in requirements engineering to resolve conflicts in stakeholders' goals. Their findings that arguments lead to consistent responses aligns with our observation. Williams and Williamson [157] incorporate arguments and Bayesian networks for breast cancer prognosis, where they exploit arguments to develop an explanation for the prognosis. A similar application in the privacy domain is to explain to a user, via arguments, the consequences of a particular sharing decision.

6.9 Conclusions

Sharing all kinds of information on SNSs is routine for many people. Examples of such information are a picture from the Christmas party and a tweet about your imminent trip with friends. The information shared often involves multiple users. When the privacy preferences of two or more users do not align, they should be able to negotiate so as to balance privacy and utility for each user. The related literature proposes methods based on aggregating privacy preferences. However, aggregation does not capture how people align with others' preferences. In contrast, we propose employing arguments to support preferences. We present an inference model that employs sharing context, and users' preferences and arguments to predict an optimal sharing policy in a multiuser scenario. We conduct a crowdsourcing study to train and test our model.

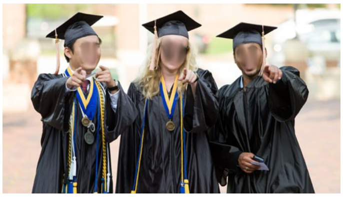
Via a series of multinomial logistic regression models, we show that all three feature types we consider influence the optimal sharing policy in a multiuser scenario. We find that (1) among the contextual variables, sensitivity has the highest influence on the optimal policy; (2) among preference-based features, the most restrictive policy has the highest influence on the optimal policy and, in particular, not the majority policy; and (3) users may value arguments for sharing more than arguments for not sharing; however, if the argument for not sharing is an exceptional case argument, users usually support not sharing.

By training an inference model for each feature type, we find that a model employing argument-based features predicts optimal policy with higher accuracy than those not employing arguments. Further, from self reported data, we find that introducing arguments increases a user's confidence in choosing the final sharing policy. Further, we investigate the data instances our prediction model misclassified to find out whether some combinations of scenario-defining elements make the prediction of optimal sharing policy difficult. We find that conflicts between arguments and preferences generate the majority of misclassification. This indicates that users interpret and prioritize arguments subjectively, which suggests that a tool that aims at helping users deal with multiuser conflicts must be adaptive and learn from the user.

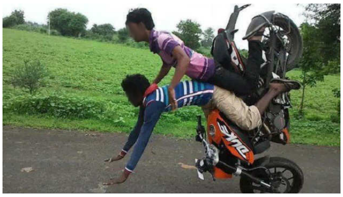
Appendix A: Picture Survey Scenarios

This Section shows the 12 pictures we employ in the picture surveys (Section 6.3.1.2). Description, context, and arguments are provided for each picture.

- The description includes the scenario in which the picture was taken and the people involved in its sharing.
- The context includes relationship, sensitivity rating, and sentiment rating of the picture. Note that the context was identified by MTurk participants who answered picture surveys corresponding to the picture.
- The arguments are of types: positive consequence, negative consequence, and exceptional case. Note that (1) each argument is employed along with a sharing policy, and (2) different combinations of arguments are employed in each picture survey (examples in Table 6.1).

Picture and Context	 Relationship: Friends (92.2%) Sensitivity rating: $\mu = 1.56$ ($\sigma = 0.96$) Sentiment rating: $\mu = 1.77$ ($\sigma = 1.46$)
Description	Tim, Ashley, and Jerry just graduated. Tim's father took the picture above after the graduation ceremony. Tim wants to upload the picture to his social media account.
Arguments	
	Positive consequence argument People we know will be happy to see that we are finally done with college. **Negative consequence argument** Our gestures are not appropriate for a moment like this; people might think that we did not take our college time seriously. **Exceptional case argument** This is not like any of our other pictures. It was our graduation, which happens only once in our lifetimes.

Picture and Context



Relationship: Friends (98.3%)

Sensitivity rating: $\mu = 3.29$ ($\sigma = 1.16$)

Sentiment rating: $\mu = 3.82$ ($\sigma = 1.11$)

Description Three friends, Santosh, Arun, and Nitin, decided to perform some stunts on a motorcycle. Unfortunately, while performing a stunt, Arun and Nitin had a minor accident. Santosh took the picture below at that very moment. Santosh wants to upload the picture to his social media account.

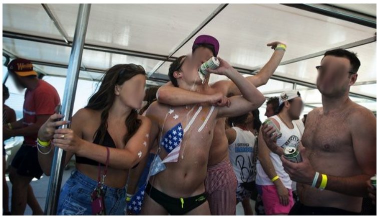
Arguments

Positive consequence argument Fortunately, none of us got hurt. This picture makes anyone who sees it laugh out loud.

Negative consequence argument People looking at this picture may think that we are reckless drivers, which is not true.

Exceptional case argument Motorbike stunts are not something we do everyday.

Picture and Context



Relationship: Friends (96.5%)

Sensitivity rating: $\mu = 3.78$ ($\sigma = 1.16$)

Sentiment rating: $\mu = 2.99$ ($\sigma = 1.26$)

Description Three friends, Mark (the groom), Alex, and John, go on a boat in Ibiza during a bachelor party. They get drunk and meet some girls. This is one of the pictures Alex took during that party. Alex wants to upload the picture to his social media account, the day after the party.

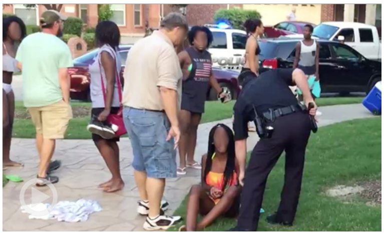
Arguments

Positive consequence argument This was one of the best day of our lives; we should share these good memories with others.

Negative consequence argument There were some girls in the party; people might understand things the wrong way.

Exceptional case argument This is not like any of our other pictures. This was Mark's bachelor party!

Picture and Context



Relationship: Friends (94.5%)

Sensitivity rating: $\mu = 4.11$ ($\sigma = 1.13$)

Sentiment rating: $\mu = 4.24$ ($\sigma = 1.20$)

Description The picture above is from an end-of-semester party Vanessa (the girl wearing dark shirt with the US flag on it) hosted. The party went wild and a neighbor called the police. There was some tension after the police arrived and a few people invited to the party were arrested. The picture also shows Vanessa's friends Natasha (the girl on the grass) and Jessica (the girl at the very left of the picture). Vanessa and her friends felt that the police abused their authority. Vanessa wants to upload the picture to her social media account, a few days after the incident.

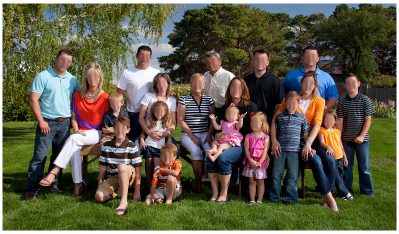
Arguments

Positive consequence argument Police arrested some of us in the party. This picture shows that police abused their power and arrested some of us for a silly reason.

Negative consequence argument We do not deserve what happened and this picture brings back those traumatic memories.

Exceptional case argument This pictures reminds us of one of the wildest party we have had.

Picture and Context



Relationship: Family (97.7%)

Sensitivity rating: $\mu = 1.90$ ($\sigma = 1.25$)

Sentiment rating: $\mu = 2.01$ ($\sigma = 1.61$)

Description The Moore brothers and their parents, wives, and children took part in a photoshoot. The following is the best picture from the photo shoot. Frank wants to upload the picture to his social media account.

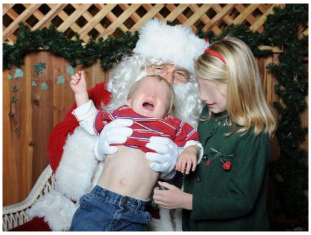
Arguments

Positive consequence argument We took this picture so that our loved ones can see it an remember us.

Negative consequence argument Our children appear in this picture; what others can do with this picture concerns me.

Exceptional case argument This is not like any other picture; we hired a photographer to take it.

Picture and Context



Relationship: Family (96.7%)

Sensitivity rating: $\mu = 2.52$ ($\sigma = 1.28$)

Sentiment rating: $\mu = 3.09$ ($\sigma = 1.06$)

Description Jerry (the grandfather dressed as Santa), Timmy (Jerry's grandson), and April (Jerry's granddaughter) took a picture during Christmas night. Jerry wants to upload the picture to his social media account.

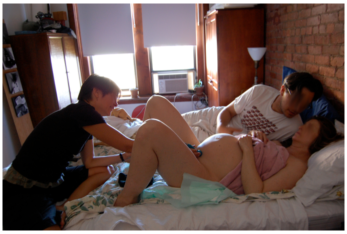
Arguments

Positive consequence argument It was a lovely evening and this picture brings back good memories.

Negative consequence argument Poor Timmy got scared. I do not think it is fair to Timmy to share a picture of him crying and scared of his grandpa.

Exceptional case argument This was the first time Jerry and Timmy took a picture together.

Picture and Context



Relationship: Family (95.3%)

Sensitivity rating: $\mu = 4.51$ ($\sigma = 0.76$)

Sentiment rating: $\mu = 2.45$ ($\sigma = 1.35$)

Description Dolores and Philip decide to have their baby, Rose, at home with the help of Ann, who is Dolores' sister and a doula. They took the picture below during the labor. Philip wants to upload the picture to his social media account, a few days after Rose was born.

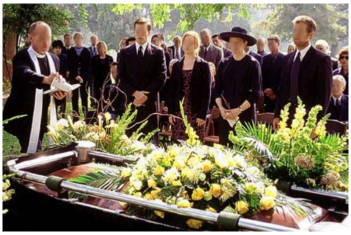
Arguments

Positive consequence argument This picture shows that Rose was born peacefully at our home surrounded by those who love us.

Negative consequence argument The idea of giving birth at home was to be at a private place surrounded by only the people we love.

Exceptional case argument This is not like any other family picture; this shows that Rose is coming to our lives.

Picture and Context



Relationship: Family (96.5%)

Sensitivity rating: $\mu = 4.21$ ($\sigma = 1.01$)

Sentiment rating: $\mu = 3.93$ ($\sigma = 1.16$)

Description Mary, Sophia, and Charles attend their mother's funeral. Another family member takes some pictures and circulates them among the family members. Mary wants to upload the picture to her social media account.

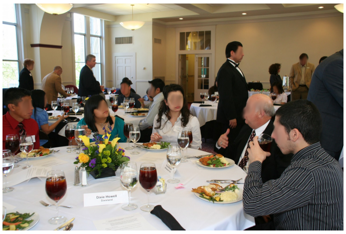
Arguments

Positive consequence argument Many people knew our mother and loved her, including our friends. When people see this picture they will remember her and see that we all were there to say goodbye.

Negative consequence argument This picture may appear highly inappropriate to many people.

Exceptional case argument This is not like any other picture, we were saying goodbye to mom.

Picture and Context



Relationship: Colleagues (94.4%)

Sensitivity rating: $\mu = 1.77$ ($\sigma = 1.10$)

Sentiment rating: $\mu = 2.83$ ($\sigma = 0.92$)

Description Maria, Bonita, and Felipe, three junior employees in a company, attend a business lunch in which they meet their seniors. One of the other employees took the following picture and sent it to Maria. Maria wants to upload the picture to her social media account.

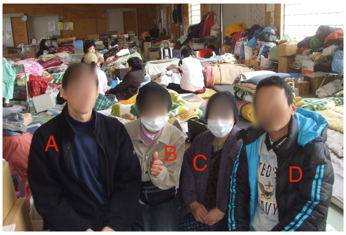
Arguments

Positive consequence argument This picture shows that we are making good progress in our careers.

Negative consequence argument This was a professional event and our seniors might want to keep it private.

Exceptional case argument This is an exceptional event since we attended a professional party for the first time.

Picture and Context



Relationship: Colleagues (86.7%)
Sensitivity rating: $\mu = 2.67$ ($\sigma = 1.32$)
Sentiment rating: $\mu = 2.69$ ($\sigma = 1.24$)

Description Aiko (C) took the picture above with her colleagues Ichiro and Katsu and, a French volunteer at the tsunami relief center. Aiko wants to upload this picture to her social media account.

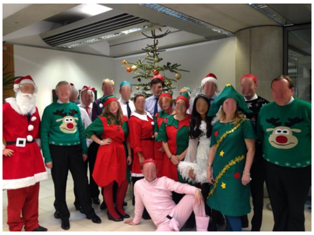
Arguments

Positive consequence argument The picture shows the difficult situation in which the survivors live. Sharing this can encourage people to help.

Negative consequence argument Tsunami was a disaster and our gestures are not appropriate. People may get the wrong idea.

Exceptional case argument This was one of the worst natural disasters.

Picture and Context



Relationship: Colleagues (92.9%)
Sensitivity rating: $\mu = 3.26$ ($\sigma = 1.41$)
Sentiment rating: $\mu = 2.46$ ($\sigma = 1.50$)

Description Jerry, Laura, and Sabrina work together in a company. They were asked to attend the Christmas party dressed. However, a guy in their company (the one in pink dress) brought the whole dressing to a new level. They took the following picture at the party. Jerry wants to upload the picture to his social media account, a few days after the party.

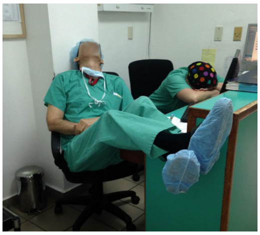
Arguments

Positive consequence argument People think that I have a boring life because I work at a boring place; this will prove them wrong.

Negative consequence argument This is embarrassing; people will pick on us because of this picture.

Exceptional case argument This is an exceptional event since a Christmas party happens only once a year.

Picture and Context



Relationship: Colleagues (98.0%)

Sensitivity rating: $\mu = 3.52$ ($\sigma = 1.26$)

Sentiment rating: $\mu = 3.50$ ($\sigma = 1.00$)

Description The hospital where Bryan, Martin, and Sophia work has recently changed its shift policy making shifts much longer. Doctors complain that these shifts leaves them exhausted. During one such long shifts, at 4am, Bryan takes a picture of his two colleagues Martin and Sophia sleeping while they wait for another patient to come to emergencies. Bryan wants to upload the picture to his social media account, a few days after the picture was taken.

Arguments

Positive consequence argument This new shift policy is too demanding. A picture like this can convince the management to change the policy.

Negative consequence argument If people think that we sleep on our job, they won't trust the hospital.

Exceptional case argument A doctor sleeping on a chair is exceptional.

CHAPTER

7

Muppet: Recommending Sharing Policies in Multiuser Privacy Scenarios

AUTHORS:
RICARD L. FOGUES[⋆], PRADEEP MURUKANNAIAH[§], JOSE M. SUCH[†], AND MUNINDAR SINGH[♠]
rilopez@dsic.upv.es, pkmvse@rit.edu, jose.such@kcl.ac.uk, mpsingh@ncsu.edu
[⋆]DEPARTAMENTO DE SISTEMAS INFORMÁTICOS Y COMPUTACIÓN
UNIVERSIDAD POLITÉCNICA DE VALENCIA, SPAIN
[§]ROCHESTER INSTITUTE OF TECHNOLOGY, NY, USA
[†]KING'S COLLEGE DEPARTMENT OF INFORMATICS LONDON, UK
[♠]DEPARTMENT OF COMPUTER SCIENCE AT NC STATE, NC, USA

Abstract

A multiuser privacy scenario involves sharing information concerning multiple users, e.g., a group photo or a post mentioning others. In current practice, one user (the sharer) sets the sharing policy in a multiuser scenario. However, this may lead to privacy violations for other users involved in the scenario. It can be challenging for the sharer to identify a policy satisfying all users involved. To address this challenge, we propose Muppet, which recommends a sharing policy considering the (1) contextual features defining the scenario, (2) characteristics of the users involved, (3) their preferences, and (4) group characteristics. Muppet eases the burden of setting an appropriate sharing policy in a multiuser scenario by reducing the user effort required to manually negotiate the appropriate policy. Muppet works incrementally and asks for users' input only when required, and it learns as users employ or discard its recommendations. However, the cold start problem typical of recommender systems is a key challenge, i.e., the issue that a system cannot draw inferences when it has not gathered sufficient information. We address this challenge by bootstrapping Muppet with crowdsourced training data, asking members of the crowd to answer questions about a variety of multiuser scenarios. We evaluate Muppet and its bootstrapping approach via an empirical study involving 988 Amazon Mechanical Turk users. We find that Muppet recommends sharing policies with higher accuracy than baseline approaches and that the quality of Muppet's recommendations increases as users provide more information.

7.1 Introduction

Sharing information on social media is commonplace. For example, taking a photo at a social event and sharing it on social media is easy and simple—thanks to the integration of social media with mobile devices. Often, the shared information involves several individuals. For example, a photo at a dinner with friends uploaded to a social network service (SNS) concerns all friends who appear in the photo.

SNSs provide facilities such as tagging (e.g., Facebook) and mentioning (e.g., Twitter) for a user to link a piece of information to other users. Although a user may enjoy the ability to link information to several users, this may present a potential privacy risk for the linked users. Suppose Alice uploads a photo from last weekend's party in which she appears with her friend Bob to an SNS and tags Bob. Bob may find that the photo Alice uploaded is sensitive. However, Bob has no control over uploading that photo, and Alice's action threatens Bob's privacy by revealing sensitive information about him from one setting (or context) into another. We identify a situation such as this as a *multiuser privacy scenario* or, for brevity, a *multiuser scenario*.

Currently, SNSs do not provide mechanisms to handle multiuser scenarios [42]. Instead, the uploader is in charge of setting the sharing policy. If the sharing policy chosen by the uploader is not appropriate for any of the other users, their privacy may be at risk. To cope with this problem, a user may employ strategies [159] such as self censorship, untagging themselves, and blocking users. However, these strategies may be ineffective in dealing with multiuser privacy. For example, it can take a long time for a user to realize that an inappropriate photo involving him is being shared.

Our key objective is to build a recommender for multiuser scenarios that recommends a sharing policy, requiring little input from the users. To this end, we (1) identify important factors that potentially influence the inference of sharing decisions in multiuser scenarios; (2) develop Muppet, a recommender system that asks for user input, incrementally; and (3) evaluate the quality of Muppet's recommendations.

Researchers, e.g., [146, 159], have identified the lack of decision-support systems for resolving multiuser privacy conflicts as a major gap in privacy management for social media. However, they do not provide a solution, as we do. Others [20, 143, 147] have proposed methods to automatically determine solutions to conflicts based on users' privacy preferences. These methods suffer from two main limitations. First, they either aggregate preferences in a fixed way, regardless of context, or consider only a few, predetermined situations among the large number of potential situations. Second, they do not consider a user's characteristics such as demographics and social media practices. However, evidence suggests that demographics, such as gender and age, influence how users share on social media.

Contributions

1. We propose Muppet, a recommender of sharing policies for multiuser scenarios. Muppet (machine) learns to recommend a sharing policy from four types of features: contextual factors, user preferences, user characteristics, and group characteristics.

2. We describe an approach to bootstrap Muppet, addressing the cold start problem by building a training dataset via crowdsourcing.

3. We evaluate Muppet on data collected from a human-subject study. We find that (1) the quality of Muppet's recommendations is better than those of a number of baseline and preference-aggregation approaches, (2) Muppet's recommendations improve incrementally, as more information is provided, and (3) Muppet provides recommendations even when users do not specify all the information.

7.2 Muppet Overview

Figure 7.1 shows Muppet's overview. Given a multiuser scenario, Muppet recommends a sharing policy, e.g., *share with common friends*. Muppet operates in three rounds and makes up to three recommendations, incrementally, for a multiuser scenario—up to one recommendation in each round. Muppet's goal is to recommend a sharing policy suitable for all users involved in a scenario. Thus, in each round, Muppet makes the same recommendation for each user involved. Muppet's recommendations in each round are independent of its recommendations in previous rounds.

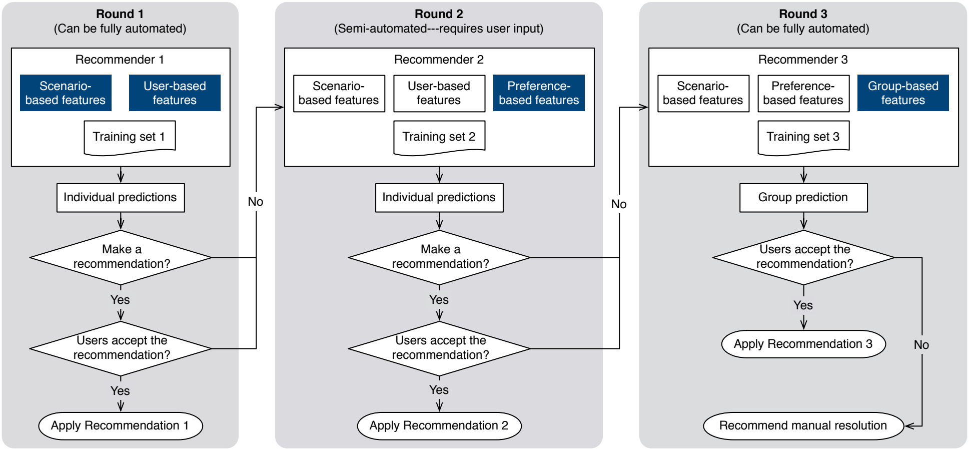


Figure 7.1: Overview of Muppet's incremental recommendation approach, highlighting the information added in each round.

In the first or second round, considering the available information, Muppet may not be able to recommend a sharing policy suitable for all users involved. In such cases, Muppet proceeds to the following round, automatically. In the third round, Muppet always makes a recommendation. During the first and second round, the users involved may accept the policy Muppet recommends, in which case Muppet stops. If not all users accept Muppet's recommendation, it proceeds to the next round. If Muppet does not achieve unanimous acceptance in the third round, it gives up and asks users to proceed manually.

Organization

Three key elements in Muppet's design are its *features*, *recommenders*, and *training data*. We provide details of these in Sections 7.3, 7.4, and 7.5, respectively. Section 7.6 describes our evaluation and Section 7.7 discusses the results. We describe related works in Section 7.8 and conclude in Section 7.9.

7.3 Features: Context, Users, Preferences, Group

Muppet employs four categories of features across the three rounds: (1) contextual; (2) individual users' characteristics; (3) users' preferences; and (4) users' group characteristics. Our choices of features in each of these categories are motivated by information easily available on current social media services, existing literature on factors influencing privacy decisions, and our intuitions.

7.3.1 Contextual Features

Our definition of context reflects intuitions compatible with those of Nissenbaum's [109] study of social settings that dictate the flow of personal information. Her definition of contextual integrity captures the idea that people share information of varying type and sensitivity in society, not simply as individuals in an undifferentiated social world, but as individuals in certain capacities (roles), in distinctive social contexts, such as health care, education, and employment.

Relationships among the individuals The roles of the individuals involved in a transaction are a defining element of context. Also, the relationship type is crucial when making *individual* decisions about privacy in social media [35]. Specifically, people share information differently with friends, family, and colleagues.

Sensitivity of the information A principle of contextual integrity is that individuals are entitled to protect their intimacy and any information they deem sensitive. The sensitivity of the information influences individual sharing decisions in social media [35].

Sentiment of the information It is known that users employ social media as a means to build social capital [141]. When building this capital, it is important to consider the sentiment that the disclosed personal information conveys. For example, a user may share a picture of his recently broken leg with friends to receive emotional support.

7.3.2 User Characteristics

Muppet exploits research on privacy behavior on social media that shows that demographics, social media practices, and sharing behaviors influence how and what information is disclosed.

Gender: Research on social media shows [140] that women tend to be more conservative than men in sharing information on social media.

Age: Research shows [156] that young social media users tend to share more freely than old users.

Education Level: This is usually available to SNSs, because they often need it to recommend friends and relationships.

Use frequency: Daily users are expected to share more information and tailor how it is disclosed.

Sharing experience: Users who have shared information on social media would have learned the implications of their privacy settings.

Conflict experience: Dealing with previous multiuser scenarios can affect the way users think about new scenarios and how they can solve them.

7.3.3 Preferences

Considering the different combinations of preferences that a multiuser scenario may present, we employ the following preference-based features.

Most restrictive policy selects the policy that restricts the audience the most. For instance, in a scenario where the preferences are *share with common friends* and *share with all*, the most restrictive is *share with common friends*.

Least restrictive policy is the opposite of the *most restrictive policy*.

Majority policy selects the policy that is preferred by the majority of the involved users. (It handles ties as well)

7.3.4 Group Characteristics

To represent the *group* of individuals involved in a particular multiuser scenario, we employ descriptive statistics based on the individual user characteristics presented previously. The statistics are the obvious ones of *Mean*, *Maximum*, *Minimum*, *Standard deviation*, and Δ, which is the range of Maximum minus Minimum. These statistics describe all characteristics but gender, which can take one of the following values: (1) all men, (2) majority men, (3) all women, and (4) majority women.

7.4 Recommenders

Muppet aims to provide personalized recommendations without requiring extensive user input in the process. To this end, Muppet follows an incremental recommendation approach divided into three rounds. Thus, although Muppet exploits a variety of features described above, it does not require all features in each round.

7.4.1 First Round

This round can be fully automated so recommendations can be provided quickly. In this round, Muppet employs only context and users characteristics as input, and as explained below, these can be obtained without user participation.

Users characteristics are easily available in an SNS, which usually track the frequency of use, as well as how often users share information. Similarly, the SNS can monitor how many conflicts have been previously resolved through Muppet—to determine the conflict experience of each party.

Regarding contextual features, first, Muppet tries to infer these automatically. To this end, Muppet can employ approaches, e.g., [41, 113, 165], to automatically infer relationship, sentiment, and sensitivity. If Muppet is unable to infer any of these features, it asks the users about them, who may not necessarily be able to provide their values. In that case, Muppet generates recommendations even when some of the contextual features are unknown.

Recommendations in this round are made on an individual basis. Once presented to the users, they may accept these recommendations or choose whatever they prefer. Regardless of their decisions, if all users agree on a sharing policy, the recommendation process completes. Otherwise, Muppet moves to the next round.

7.4.2 Second Round

Here, Muppet employs context, user characteristics, and the preferred sharing policies of the users expressed in the previous round. In particular, it employs the features as used in Round 1 (context and user characteristics), and the features that represent the preferences expressed in Round 1: most restrictive policy, least restrictive policy, and majority policy.

As before, Muppet suggests a sharing policy individually. If all agree, that sharing policy is applied. Otherwise, Muppet moves to the last round of recommendation.

7.4.3 Third Round

The main difference with the previous rounds is that Muppet generates a recommendation for the whole group of users. To do this, Muppet employs a model that does not consider individual characteristics, instead it employs features that define the group. The recommendation is either accepted by all users or dismissed if anyone vetos it. In the latter case, Muppet gives up and lets the users proceed manually. However, Muppet stores the final decision, if any, for further learning.

7.5 Bootstrapping Muppet via Crowdsourcing

Muppet learns from data on how groups of users make sharing decisions in a variety of multiuser scenarios. That is, given historical data on multiuser scenarios in which a user was involved and sharing decisions employed in those scenarios, Muppet can learn to recommend a sharing policy for a new scenario for that user. But Muppet must overcome the *cold start* problem: collect data about multiuser scenarios a user encounters before making recommendations.

Since there may be a large variety of multiuser scenarios, collecting sufficient data from users would be time consuming. Instead, we seek to bootstrap Muppet by *crowdsourcing* a dataset consisting of a variety of multiuser scenarios and a suitable sharing policy in each of those scenarios. Then, this dataset can be used to train Muppet before putting it to use so that Muppet can make recommendations starting from the first multiuser scenario in which a user employs Muppet.

Building a training dataset via crowdsourcing enables collecting data from a large number of users of diverse backgrounds in a short amount of time and fairly inexpensively. However, a key challenge is to design *microtasks* to obtain useful training data from the crowd.

We expect that, in a multiuser scenario, one user's preferred sharing policy will rarely match everyone's preferences. However, it is difficult to recruit cliques of participants that already know each other and ask them to recreate multiuser scenarios. Also, users are often reluctant to share sensitive information (one of the contextual factors in our model), biasing the study toward nonsensitive cases that users are willing to reveal [150]. An alternative is asking users to self-report how they behave when they experience a multiuser scenario, but the results may not match participants' actual behavior because of the well-known dichotomy between users' stated privacy attitudes and their actual behavior [1].

Considering the challenges above, we chose to create *situations* in which participants are *immersed* [100] to improve behavior elicitation while avoiding biasing the study to nonsensitive situations. We present information about two or more individuals in a specific circumstance: a combination of context and preferences. We ask participants to choose a suitable sharing policy for that circumstance.

We recruited participants for our study from Amazon MTurk [110]. We directed each participant to an external website that asked the participant to complete seven survey instruments: a presurvey questionnaire about demographics, five picture surveys (each involving a privacy conflict scenario and two sets of questionnaires), and a post-survey questionnaire about the participant's general opinions about resolving multiuser privacy conflicts. We obtained IRB approval for our study.

7.5.1 Presurvey Questionnaire

We asked participants to report their age, gender, level of education, how frequently they use social media, and how often they share (multiuser) pictures online. Since some of the situations we presented could be inappropriate for young readers, we required participants to be older than 18 years of age and showed a disclaimer at the beginning that the survey may be inappropriate for some users.

7.5.2 Picture Survey

The picture survey is the core of our study. We show a picture and describe a hypothetical scenario in which the picture was taken and next ask two questionnaires (Q1 and Q2). Table 7.1 shows one example of picture survey. We generated this

Picture	
Description	Aiko (C) took the picture above with her colleagues Ichiro and Katsu and, a French volunteer at the tsunami relief center
Rating	Identify the relationship between Aiko, Ichiro, and Katsu and rate the sensitivity and sentiment of the picture
Context (Q1)	Consider that Aiko wants to upload this picture to her social media account. What sharing policy should she apply for the picture?
Preferences (Q2)	Next, consider users' preferences as follows **Aiko** Share among ourselves **Ichiro** Share among ourselves **Katsu** Share with all Considering the context and users' preferences, what sharing policy should Aiko apply for the picture?

Table 7.1: Shortened example of a picture survey.

and several similar picture surveys by combining factors identified in the previous Section, as described below.

1. Regarding context variables, we consider a predefined set of relationship types, namely, *friends*, *family*, and *colleagues*. Related research works, such as [148], employ the same three types as an approximation to a user's relationships on a social network. Further, we assume that all individuals involved in a scenario have the same type of relationship with each other (i.e., all are either *friends*, *family*, or *colleagues*). Also, the pictures shown in the situations could be sensitive or nonsensitive and convey a positive or a negative sentiment. This leads to 12 possible contexts.

   Although we selected 12 representative pictures, one for each combination, it is important to note that we ask participants to identify the contextual factors for each picture shown to them as indicated in Table 7.1.

   Specifically, we ask participants to identify the type of relationship among the people involved in the scenario (family, colleagues, or friends), and rate sensitivity (Likert scale 1 = not sensitive at all, 5 = very sensitive), and sentiment (1 = extremely positive, 5 = extremely negative). It is worth noting that participants agreed with our assessments more than 80% of the time.

2. A sharing policy can imply no sharing, sharing publicly, or anything in between. Further, depending on the number of contacts and their types, sharing policies change from one SNS user to another. The space of possibilities is large. For simplicity, to bootstrap Muppet, we consider only three levels of disclosure:

   (a) *Share with all*: Anyone on the SNS can access the information.

   (b) *Share among themselves*: Only the individuals directly connected with the information can access the information. Since the scenarios presented in our study always include a number of individuals who are members of a group picture, the case of sharing among themselves equals the case of no sharing. That is, it does not matter whether the picture is shared among themselves on the

SNS, because they shared the moment when the picture was taken, and sharing the moment can be more meaningful than sharing it online. Consequently, this preference is the direct opposite of *share with all*.

(c) *Share with common friends*: Only common friends of the individuals involved in the scenario can access the information. As explained before, the space of possibilities is large and varies from user to user. Thus, no predefined privacy preferences cover every possibility. Nonetheless, we assume that *share with common friends* is a reasonable compromise to select this option because it lies in between the two ends of the spectrum represented by the preceding two preferences.

3. We limit the number of individuals involved in each scenario to three. This way, scenarios could present a policy chosen by a majority without ties. Although some pictures showed more than three individuals, our scenarios discussed the preferences of only three individuals.

4. We make sure that not all three individuals in a scenario use the same policy preference. Our objective is to understand how a user decides a final policy given the scenario and the preferences of others in the scenario. If all users thought the same way and wanted the same result, the solution would be trivial.

Putting the above together, we have: 12 pictures based on context, three policy preferences the first two individuals can employ, and two preferences the last individual can employ (last restriction above). That is, we generated 216 scenarios. Each MTurk participant was shown five unique scenarios, making sure that no participant was shown the same picture twice. Scenarios were randomized to counter ordering bias. Further, we asked participants to immerse themselves in the particular scenario and ignore the resemblance or lack of resemblance of each scenario to other scenarios in which they might have seen that picture.

Following the picture and its description, we asked participants to identify the contextual factors for the scenario and answer two sets of questionnaires (Q1 and

Q2). We asked participants to answer these questionnaires sequentially and when answering a questionnaire, to consider only information provided to them up to that point.

Each of the two questionnaires tells participants that one of the individuals in the scenario wants to upload the picture to a social media account and asks participants what sharing policy should be applied. The participants choose one of the policies from *share with all*, *share with common friends*, and *share among themselves*. In the first questionnaire (Q1), participants know only the contextual attributes, but not the preferences of the individuals in the scenario. This case is similar to a real scenario where a user wants to upload and share information without asking others potentially concerned with the information. The second questionnaire (Q2) introduces the preferences of all the users.

7.5.3 Participants and Quality Control

The number of unique participants that completed the study was 988. This guaranteed that each scenario had received at least two responses. Compensation was provided for only those who completed all seven steps in the survey.

For quality control, we required participants to have completed at least 50 tasks on MTurk and to have had a success rate of at least 90% [114]. We included an attention check question [50] in the ratings section of each picture survey, asking how many people (faces) were present in the picture, answering which requires counting from the picture. Participants answered the attention question incorrectly in a total of 38 instances (less than 1% of responses). If a participant incorrectly answered the attention check question in a picture survey, we excluded that picture survey from analysis, and retained only those picture surveys where the participant answered the attention check questions correctly.

Table 7.2 summarizes our participants' responses to the presurvey questionnaire.

Gender	Male: 46.3%, Female: 53.4%, Other: 0.3%
Age	18–20: 2%, 21–29: 36.6%, 30–39: 36%, 40–49: 13.7%, 50–59: 7.5%, 60 or more: 4.1%
Education	Graduate degree: 11.2%, Bachelor degree: 44.4%, College no degree: 30.9%, High school: 12.4%, Less than high school: 1%
Social media usage	Daily: 83.9%, Weekly: 12%, Monthly: 3.7%, Never: 0.4%
Pictures shared	Many (>5): 35.1%, Few (1–5): 45%, None: 18.1%, Not sure: 1.7%
Conflicts experienced	Many (>5): 2.8%, Few (1–5): 30.1%, None: 66%, Not sure: 1.1%

Table 7.2: Demographics of MTurk participants of our study.

7.5.4 Training Muppet's Classifiers

As shown in Figure 7.1, Muppet employs a machine learning classifier in each round. We train these classifiers using data collected in the study. We gathered 3,767 valid responses for photo surveys, of which we use 70% (2,637) to build training datasets, and the other 30% (1,130) for evaluation. All the machine learning classifiers were implemented using Weka [60]. The specific choice of machine learning classifier was made based on results empirically obtained with the data collected in the study.

To build the training set employed in Round 1, we use the responses given by the participants in the first questionnaire (Q1) of the photo surveys. Each response is a sample, and its features are the ratings of the contextual elements and the characteristics of the participant that provided that response. The class of the sample is determined by the sharing policy chosen by the participant. With this dataset, we train a random forest classifier [17].

To build the training set for Round 2 classifier, we use the responses given by the participants during the second questionnaire (Q2) of the photo survey. Again, each response is a sample with its features as above. However, this time, we include preference features, specified based on the preferences presented in the scenario associated with the response. For example, if the scenario of a given response presented two users preferring *share with all* and one preferring *share with themselves*, the feature *majority policy* would be *share with all*. As in Round 1, the class of each sample is determined by the sharing policy chosen in the response. With this dataset we train a logistic regression classifier.

Finally, for the third round, we build a training set employing each possible triplet of responses provided for the Q2 questionnaire. To form a triplet, the three responses must share the contextual and preference features. We consider each triplet as a unique sample with the following features: contextual, preferences, and group characteristics. The class of each sample is determined by the majority policy in the response triplet, or *share with common friends* in case of a tie. For example, if a triplet is formed by two responses that chose *share with all* and one that chose *share with common friends*, the class of the triplet is *share with all*. Using this dataset, we train a random forest classifier.

7.6 Evaluation of Muppet Bootstrapping

We describe our evaluation techniques and results.

7.6.1 Evaluation Strategy

To generate the testing dataset, we form triplets of the 1,130 responses we set aside for testing. All responses in a triplet must share the same contextual and preference features.

After all possible triplets are found, we calculate the accuracy yielded by Muppet for each one. To calculate the accuracy, we run the triplet through Muppet as it was an actual multiuser scenario. After each round, if a recommendation is generated, we compare the recommended sharing policy with the actual sharing policy of each response in the triplet. If both are equal, Muppet recommended a sharing policy correctly, otherwise, it failed. We normalize the accuracy obtained for each triplet: hence, the possible accuracy values are 0, 0.33, 0.67, and 1.

To compare the results obtained by Muppet, first, we employ three baseline approaches, *Always Self*, *Always Common*, and *Always All*, based on always recommending the respective sharing policy. Second, we employ recommenders based on preference aggregation: *veto* and *majority*. *Veto* recommends the most restrictive policy among those preferred by the users. *Majority* recommends the policy preferred by the majority of the users.

We show the results as bar plots. According to Lilliefors test [91], the results obtained do not come from a normally distributed population. Therefore, to test the significance of the results, we employ Kruskal-Wallis test [31], a nonparametric test that does not assume a normal distribution. Finally, in the evaluation we make several comparisons, to minimize Type I errors (false positives), we apply the Holm-Bonferroni correction [67] during the evaluation of the statistical significance of the results.

7.6.2 Hypotheses

We evaluate the following hypotheses.

- *H-Baseline*: After bootstrapping Muppet, it achieves a better accuracy than the baseline recommenders.

- *H-Aggregation*: After bootstrapping Muppet, it achieves a better accuracy than recommenders based on preference aggregation.

- *H-Incremental-Improvement*: The quality of recommendations increases as additional information is specified.

- *H-Incremental-Generation*: The probability of generating a recommendation before reaching the third round increases as additional contextual features are specified.

7.6.3 *H-Baseline*

From the 1,130 responses of the training set, we can form 510 suitable triplets. Figure 7.2 shows the accuracy results obtained by Muppet and the baseline recommenders. All the differences in accuracy are significant with a 95% confidence.

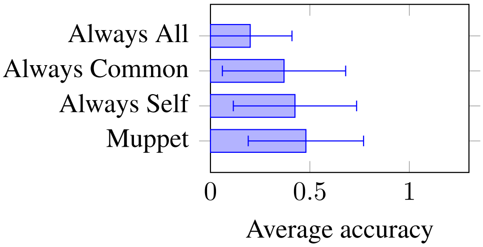


Figure 7.2: Comparison with baseline recommenders (p-value<0.01).

The main goal of the study design was to collect data for bootstrapping Muppet. Therefore, it does not contain information about how users would interact with Muppet. Specifically, we do not have data about how users could accommodate Muppet's recommendations or reach an agreement about the optimal sharing policy. Instead, we have several scenarios where the ground truth is in conflict. For example, the 510 triplets employed in the previous analysis contain scenarios where two participants chose *share with all* and the third chose *share with common friends*. From the test data, we can generate 71 triplets with non-conflicting ground truth (i.e., all participants chose the same sharing policy). Figure 7.3 shows the results obtained by Muppet and baseline recommenders employing these triplets. Muppet yields an accuracy close to 1, which indicates that, if an agreement between all the parties exists, Muppet is able, most of the times, to recommend the exact sharing policy that all the users would agree on. All the differences in the accuracy, except *always self* vs *always common*, are significant with a 99% confidence.

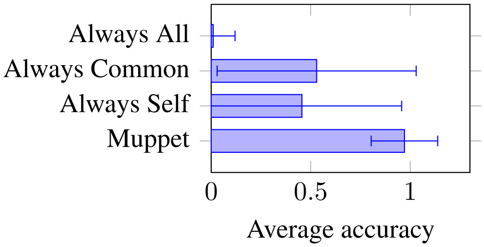


Figure 7.3: Accuracy results for scenarios with non-conflicting ground truth.

7.6.4 *H-Aggregation*

Veto and *majority* recommenders depend on the preferences of the users. Therefore, these cannot provide any recommendation in the first round of Muppet; when only contextual factors are known. To compare Muppet and these two recommenders on equal terms, we test this hypothesis employing a modified version of Muppet that never provides a recommendation in the Round 1. Figure 7.4 shows the results obtained by this version of Muppet, *veto*, and *majority*. The differences between the results are statistically significant at 95% confidence level.

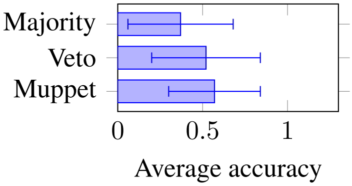


Figure 7.4: Accuracy comparison with preference-aggregation recommenders.

To refine the test of this hypothesis, Figure 7.5 shows the results obtained by these recommenders when employing scenarios with non-conflicting ground truth. Again the differences are statistically significant. The differences between the results are statistically significant at 95% confidence level.

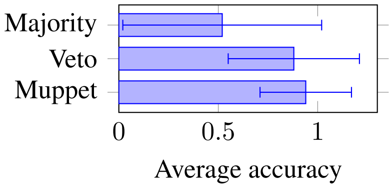


Figure 7.5: Accuracy comparison with preference-aggregation recommenders employing only scenarios with non-conflicting ground truth.

7.6.5 *H-Incremental-Improvement*

To test this hypothesis, first we look into the effects of each contextual feature on the quality of the recommendations. Figure 7.6 shows the accuracy obtained by Muppet employing one contextual feature at a time. Although the differences in average accuracies are statistically significant with a 95% confidence, they are minimal.

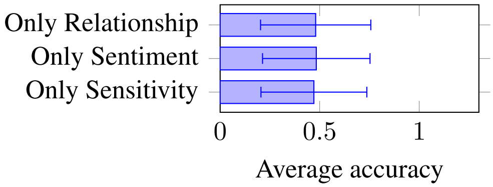


Figure 7.6: Accuracy with single contextual features.

To investigate further how each contextual feature affects the performance of Muppet, we analyze the accuracy of each individual contextual feature employing only scenarios with consensus. Figure 7.7 shows the results obtained. This analysis shows greater differences in accuracy. This time, the differences in the results are significant with 99% confidence level.

When Muppet is not able to generate a recommendation in the first round, it moves to the second and asks users for their preferences. Therefore, the amount of available information increases from Round 1 to Round 2. To test if this increase of specified information improves the accuracy of Muppet, we study the performance of each round separately. That is, when Muppet generates a recommendation, we compute

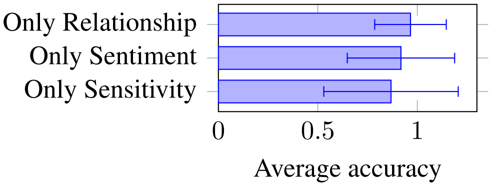


Figure 7.7: Accuracy results employing only one contextual feature at a time in scenarios with non-conflicting ground truth.

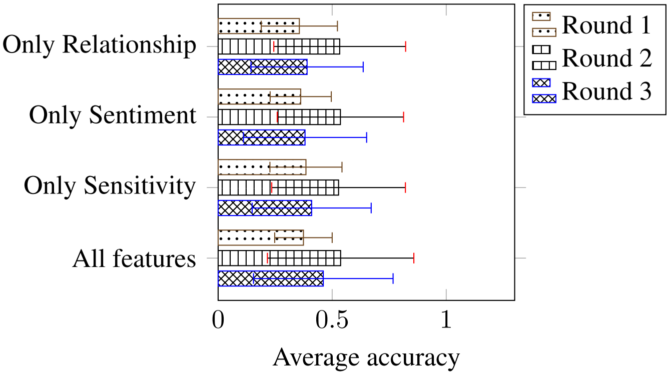


Figure 7.8: Accuracy per round.

the accuracy and assign that accuracy score to the round where the recommendation was created. Figure 7.8 shows the average accuracy achieved at each round for employing all contextual features and only one at a time. The results show that Round 2 is the most accurate overall. All the differences are significant with 95% confidence level except the results obtained by *only sentiment* and *only relationship* in Round 3.

Figure 7.9 shows the accuracy results of each round using only scenarios with non-conflicting ground truth. When employing these scenarios, we find that no recommendation is generated in Round 1. Thus, the figure does not show results for Round 1. On the one hand, the differences in the results obtained for Round 2 are not significant. On the other hand, the results obtained for Round 3 are significant with 99% confidence level, excluding *only sentiment* vs *only relationship*. As shown

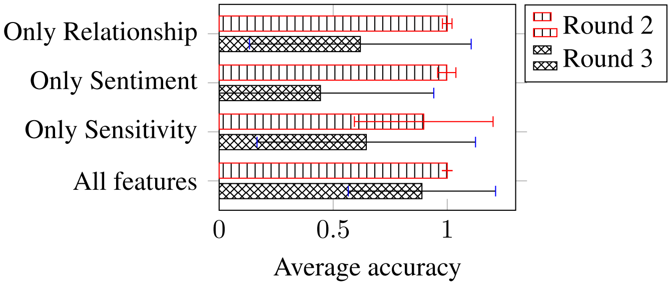


Figure 7.9: Accuracy results per round in scenarios with non-conflicting ground truth.

Evaluation	**Round 1**	**Round 2**	**Round 3**
All Features	28.24%	49.8%	21.96%
No-Conflict All Features	0%	74.29%	25.71%
Only Sensitivity	25.45%	57.51%	15.04%
No-Conflict Only Sensitivity	0%	88.6%	11.4%
Only Sentiment	16.34%	68.02%	15.64%
No-Conflict Only Sentiment	0%	85.65%	14.35%
Only Relationship	14.31%	66.9%	18.79%
No-Conflict Only Relationship	0%	91.27%	8.73%
Average	10.79%	72.76%	16.45%

Table 7.3: Recommendations generated in each round.

in the Figure, Round 2 achieves perfect accuracy in all cases but *Only Sensitivity*. *All features* achieves the highest accuracy in Round 3. Therefore, there are more chances of generating a correct recommendation in scenarios that reach Round 3 when using all contextual features than when using only one.

7.6.6 *H-Incremental-Generation*

Now, we move on analyzing the number of recommendations generated at each round depending on the number of specified contextual features. Table 7.3 shows a summary of the number of recommendations created at each round for the three types of evaluations (with all contextual features, only scenarios with non-conflicting ground truth, and with only one contextual feature). As shown in the Table, round 2 is where the majority of scenarios receive a recommendation. The results show that when employing all contextual features, the probability of getting a recommendation before specifying preferences increase. Moreover, sensitivity is the contextual feature that increases this probability the most.

7.7 Discussion of the Results

The results show that Muppet outperforms the baseline and preference-aggregation approaches. Although, the differences are statistically significant, Muppet achieves a modest accuracy. However, when we evaluate only scenarios where an agreement among all the parties is possible, Muppet is able to find the correct sharing policy that all the parties would agree on more than 97% of the time.

To test *H-Incremental-Improvement* we perform two analyses. First, we study the effect of each contextual feature when employed individually. We find that relationship is the feature that improves the accuracy the most, followed by sentiment, and lastly sensitivity. Therefore, in case contextual features cannot be inferred automatically and considering that users may stop specifying contextual features at any time, Muppet should ask users for relationship first, then sentiment, and lastly sensitivity.

Second, we study the accuracy of the recommendations generated at each round. This analysis points out that, when preferences are specified (Round 2), the accuracy increases. Furthermore, contextual features impact the accuracy of each round in two ways. Relationship and sentiment improve the accuracy of Round 2, whereas sensitivity improves the accuracy of Round 3.

Finally, we study the number of recommendations generated at each round. Most of the recommendations are provided in the second round. However, the analysis shows that when sensitivity is specified, the number of recommendations generated at the first round increases. Therefore, using this contextual feature can increase the probability of receiving a recommendation earlier in the process.

7.7.1 Future Work

This paper is a stepping stone on the Muppet development. With the bootstrapping of Muppet finished, the next step is to develop a prototype and test it with users. A prototype would enable us to assess quantitatively and qualitatively how acceptable the recommendations yielded by Muppet are.

The evaluation technique employed in this paper aims at analyzing the validity of the presented bootstrapping method. However, it does not measure the level of approval of a recommendation. For example, a user whose preferred sharing policy is *share with all* could consider as acceptable the recommendation *share with common friends*. Moreover, if, in some cases, users are willing to accommodate the recommendations provided by Muppet, it could be expected that its accuracy increases.

Although we present a bootstrapping technique that allows Muppet to avoid the usual problem of recommenders known as cold start, we envision a system that is able to learn from users and adapt to them. To that aim, we plan to develop and evaluate Muppet's adaptive capabilities.

7.7.2 Limitations

To limit the sheer number of scenarios in the study, we employed three predefined sharing policies and three contextual factors. Thus, our findings are specific to these values. Nonetheless, other relationship types, contextual factors, and sharing preferences can be appropriate or even desirable in a multiuser scenario. It is worth noting that adding scenario-defining features will increase the amount of data requested from users. Therefore, future work will consider the trade-off between accuracy and user effort.

During the evaluation of Muppet, we treat overshares (e.g., recommending *share with all* when the user would prefer *share with common friends*) and undershares (e.g., recommending *share with common friends* when the user would prefer *share with all*) as equal. However, in some cases, an oversharing recommendation can be considered worse than an undersharing one, and vice versa.

Our study is cross-sectional (one time) and in the scenarios of the study, decisions are to be made in one round. Thus, our study does not provide data to understand how users would deal with the recommendations provided by Muppet and how they could accommodate their preferences to those recommendations.

7.8 Related Work

Currently, SNSs do not offer tools that help users deal with multiuser conflicts. Owners of the information are in charge of setting appropriate sharing policies. [159] found that users rely on strategies such as unfriending, removing tags from photos, and self censorship, though these approaches have limited effectiveness in practice.

A number of research works propose tools that help users choose an appropriate sharing policy in multiuser scenarios, but these tools require intense human intervention which may overload users. [6] present a method where the owner of a picture in a multiuser scenario is in charge of deciding the sharing policy and the other users involved can suggest privacy preferences. [158] propose a method where all the parties involved in the conflict refine a sharing policy iteratively.

Some approaches provide automated support by means of preference aggregation. We can find proposals that consider only one fixed way of preference aggregation [20, 147]. Others, such as [69], consider more than one method of aggregation, but the user who uploads the information chooses the method, which becomes a unilateral decision without considering any input from others. [143] provide an improvement over the fixed ways of aggregating user preferences by automatically inferring the particular situation for the conflict and applying the *concessions* that arise during offline negotiations in those situations [159]. However, these proposals consider only preferences whereas Muppet incorporates elements such as the context of the photo being shared, the characteristics of the users, and the relationship among them. Evidence supports that these elements play a key role when users decide what to share and with whom [156].

Recently, game-theoretic mechanisms have been suggested as a means to help users manage multiuser scenarios. These research works define negotiation protocols, which are a means of standardizing the communication between users in the process of negotiating a solution to a multiuser privacy conflict by defining how users can interact with each other [144]. However, such proposals may not work well in practice since they assume users are perfectly rational and do not capture the social idiosyncrasies that users consider in real life [159].

7.9 Conclusions

Sharing photos of groups is common for many users. For example, a picture taken during a trip with friends or in a birthday party often involves many users. While sometimes finding the appropriate sharing policy is trivial, often preferences can conflict. In these cases, an inappropriate sharing policy may pose a privacy violation for some of the parties involved. Currently, SNSs leave the responsibility of finding a suitable sharing policy on the user who uploads the photo. However, users are not always able to solve the conflict or they lack the time to do so.

This paper introduces Muppet, a sharing policy recommender for multiuser scenarios that aims at providing recommendations with low user effort. To achieve this goal, Muppet works incrementally and divides the recommendation process into three rounds. To create the recommendations, Muppet employs machine learning classifiers that exploit a variety of features: contextual factors, user preferences, user demographics, and group characteristics.

The results show that Muppet outperforms baselines approaches. Moreover, we find that when there is a consensus among the users about the sharing policy, Muppet achieves a high accuracy. We also find that Muppet achieves acceptable accuracy scores when only one contextual feature is known. Finally, the analysis points out that when only sensitivity is specified, Muppet yields lower accuracy, but it is more likely to obtain a recommendation in the first round.

Part III

Discussion

CHAPTER

8

Discussion and Future Work

In this chapter, the results obtained by the contributions of this thesis are discussed, as well as possible paths for future research. Section 8.1 discusses the open challenges found during the review of research on access controls for SNSs and how this thesis addresses some of them. Experimental results obtained by Facebook users employing BFF are commented in Section 8.2. Section 8.3 analyzes the viability of tie strength and tags as new attributes for access controls on SNSs and discusses the results obtained experimentally by three ReBAC models that employ these two new attributes. Section 8.4 discusses the main aspects of modeling multiuser scenarios employing context, preferences, and arguments. Furthermore, it also analyzes the results obtained by an inference model that is able to predict the optimal policy for a given multiuser scenario, and the influence of each element on the optimal privacy policy. Finally, Section 8.5 summarizes the results obtained by the bootstrapping method employed to avoid the cold start problem of Muppet.

8.1 Results on the Definition of Open Challenges

Chapter 2 presents a thorough review about current research on access controls for SNSs. To facilitate the comprehension of how these works are linked between them, the reviewed papers and articles are classified and organized into five categories. Furthermore, the review shows conceptual maps that help readers visualize the connections among current research.

The contributions of this PhD thesis focus on addressing three of the challenges found in the review: (1) tie strength automatic inference, (2) ReBAC and content type, and, (3) multiuser privacy management. However, others remain open. For example, privacy management based on self-presentation. One of the reasons users employ SNSs is because they can build social capital. To improve the way users build social capital on SNSs, they should be able to define how they disclose information based on what image of themselves they want to project on others. For example, users trying to find a job would like to appear as good professionals to potential employers while sharing funny and entertaining content with friends. Future work should investigate new attributes and mechanisms for access controls that enable this functionality.

8.2 Results on BFF

Chapter 3 presents BFF and an empirical study to test its accuracy in tie strength and community prediction. Thirty-eight Facebook users participated in the study. The participants employed BFF to automatically group their contacts and assign a tie strength value to each one of them. Then, participants could correct any misclassifications or errors.

The results of the study indicate that BFF can alleviate the burden of specifying tie strength values on SNSs and it enables access controls to employ this attribute without increasing user effort. Additionally, BFF can help users to group their contact, which speeds up the whole privacy management process. BFF employs a tie strength predicting model that uses a reduced number of variables that are commonly available on SNSs and are not costly to collect. Although BFF was tested on Facebook, many of the elements employed in the prediction model can be easily translated to other SNSs such as Twitter or Google+. For example in Twitter, Facebook *likes* and *number of common friends* can be translated to *favs* and *number of common followers* respectively. Therefore, it is reasonable to assume that BFF can help any SNS in the transition into access controls that use tie strength as an attribute.

As pointed out in Section 3.5.3.1 one of the current limitations of BFF is that it does not employ information outside the SNS. However, some users have contacts on their profiles that have a strong tie but do not interact with them on the SNS. For example, a mother and a daughter may become friends on Facebook, but their interactions occur mainly outside of the social network. Therefore, BFF is unable to predict their tie strength accurately. Future work should increase the sources of information that BFF can draw from. These can include e-mail, phone calls, and, taking advantage of GPS integrated on mobile devices, location information [8, 106].

8.3 Results on Tie Strength and Tags as Attributes for Access Controls

As explained above, the results yielded by BFF point out that introducing tie strength into ReBAC models is viable. User effort would not increase even though they could specify more characteristics of their relationships. To investigate the practical effects of tie strength on current access controls, Chapter 4 reports on a study with human participants who provided their preferred privacy policies for photos on their Facebook accounts.

Employing the data collected in the study, two new attributes for access controls, tie strength and tags, are evaluated. Specifically, fifteen access controls with different combinations of the following attributes are tested: (1) groups, (2) tie strength, (3) individual identifiers, and (4) tags. Three metrics are employed during the evaluation: (1) coverage, (2) number of rules, and (3) complexity.

Analyzing the two new attributes individually, on the one hand, the results obtained point out that access controls with the attribute tags achieve good coverage with low complexity. This shows that tags play a key role during privacy policy definition. On the other hand, the tie strength attribute does not show an impact on access controls as positive as tags. One of the reasons behind this could be that participants did not assign tie strength values depending on how much they share on the SNS but outside of it.

Based on the results of the previous study, three access control prototypes were developed. Chapter 5 reports on a study conducted to evaluate these prototypes with users. During the study, participants defined privacy policies employing two different access controls, in this fashion, the collected data enabled a direct comparison between access controls and attributes.

The results of the study show that users prefer access controls that employ tie strength and tags. Moreover, the results also point out that when defining sharing policies, users employ these two attributes extensively. Nevertheless, the result of the study also indicate that when users employ tie strength and tags they make more mistakes in terms of sharing policy correctness; specially, when combining tags and tie strength. It is worth noting, tie strength can help users reduce the privacy breach caused by incorrect sharing policies.

Based on the results obtained, it seems appropriate to include tie strength and tags in access controls for SNSs. Nonetheless, the results also indicate that users would require tools that assist them during sharing policy definition. Concretely, SNSs developers should provide users with the following functionalities: (1) automatize contact grouping and tie strength specification, (2) automatize photo tag definition,

(3) access control personalization, and (4) a sharing policy assistant that helps users visualize and simplify sharing policies. To validate the utility of tie strength and tags in the long run, future work should test whether users educated about how to use tie strength and tags make fewer mistakes than those who use them for the first time. Additionally, future research should also investigate how users refine iteratively tie strength values and photo tags as they learn how these attributes affect their sharing policies.

8.4 Results on Modeling Multiuser Privacy Scenarios

The study shown in Chapter 6 aims at discovering what elements influence the optimal privacy policy for a multiuser scenario and creating an inference model that is able to predict such policy. First, a formal model for multiuser scenarios is defined. This model employs three elements: (1) contextual factors, (2) preferences, and (3) arguments.

A series of multinomial logistic regression models show that all three element types influence the optimal sharing policy in a multiuser scenario. The findings are: (1) among the contextual variables, sensitivity has the highest influence on the optimal policy; (2) among preference-based features, the most restrictive policy has the highest influence on the optimal policy and, in particular, not the majority policy; and (3) users may value arguments for sharing more than arguments for not sharing; however, if the argument for not sharing is an exceptional case argument, users usually support not sharing.

Regarding the inference model, the study finds that a model employing argument-based features predicts optimal policy with higher accuracy than those not employing arguments. An analysis of self reported data indicates that introducing arguments increases a user's confidence in choosing the final privacy policy. Further, the instances that the inference model misclassfies are investigated to find out whether some combinations of scenario-defining elements make the prediction of optimal privacy policy difficult. The findings point out that conflicts between arguments and preferences generate the majority of misclassification. This indicates that users interpret and prioritize arguments subjectively, which suggests that a tool that aims at helping users deal with multiuser conflicts must be adaptive and learn from the user.

8.5 Results on Muppet

In Chapter 7, the privacy policy recommender Muppet is introduced. The main goal of Muppet is to provide a policy recommendation quickly and in a way that requires little user effort. To that aim, Muppet divides the recommendation process into three rounds. In each round, specific features are employed, in this fashion, Muppet collects and processes the required information incrementally. Moreover, Muppet employs data that is available in most SNSs, thus, automatizing further the process.

One of the typical issue that recommenders encounter is *cold start*. Recommenders cannot draw inferences when they have not gathered sufficient information. To address this problem, a bootstrapping method that enables Muppet to offer recommendations off-the-shelf is presented and evaluated.

Muppet is evaluated employing data collected from a human-subject study. First, the accuracy of Muppet is compared against the accuracy yielded by three baseline approaches: (1) *Always All*, (2) *Always Common*, and (3) *Always Self*. Muppet outperforms these three baseline recommenders. Second, Muppet's accuracy is compared against two preference aggregation methods (1) *Majority*, and (2) *Veto*. Again, Muppet outperforms these two approaches. It is worth noting that Muppet achieves almost perfect accuracy when there is a consensus on the optimal privacy policy among all the parties involved.

Finally, the incremental recommendation feature of Muppet is evaluated. The results indicate that *relationship* is the feature that most influences the accuracy of Muppet. However, *sensitivity* increases the probability of Muppet providing a recommendation in Round 1, thus, reducing the user effort.

The next step in the development of Muppet is to create a prototype and test it with users. This would enable a quantitative and qualitative assessment of the acceptance of Muppet's recommendations. The presented evaluation of Muppet does not measure the level of approval of a recommendation. For example, a user whose preferred sharing policy is *share with all* could consider as acceptable the recommendation *share with common friends*. Moreover, if, in some cases, users are willing to accommodate the recommendations provided by Muppet, it could be expected that its accuracy increases.

Additionally, this prototype should include adaptive capabilities. These would allow Muppet to learn from the user and improve recommendations over time. The future study with users would provide valuable data to test and evaluate different approaches of adaptability.

CHAPTER

9

Conclusions

SNSs offer functionalities that a wide number of persons enjoy. People feel that sharing photos, events, comments, and interacting with their friends at any time and anywhere are convenient and help them build social capital. However, as the functionalities and use of SNSs are increasing, so are privacy concerns. To avoid losing potential users, SNSs have to offer privacy management systems that are easy to use and help users understand clearly who can and cannot access their information on their profiles.

Several formal ReBAC models have been suggested over the recent years [23, 46]. However, a number of unresolved issues have prevented their actual implementation on commercial SNSs. The main goal of this thesis work is to close the gap between these ReBAC models and those that are currently employed on SNSs. To that aim, a study of the state of the art in the field of ReBACs revealed a number of pressing challenges that access controls must overcome before being useful for actual SNSs. This thesis tackles three of those challenges: richer relationship definitions, privacy based on content type, and multiuser privacy management.

To translate accurately privacy preferences into sharing policies, users require rich social relationship definitions. However, this can also increase the complexity and the effort required from the users to specify properly their social connections. To solve this problem, this thesis dissertation presents BFF. This tool reduces the burden of grouping and specifying the type of each social connection. BFF employs a simple inference model that only requires fourteen variables, and offers a tie strength accuracy over 80%.

Considering that tie strength elicitation is not a burden on the users, this thesis studies how the inclusion of two new attributes, tie strength and tags, affects access controls and how users interact with them. First, this thesis reports on a study about the viability of different access controls with combinations of attributes. Employing a three metrics the access controls are evaluated and the three most promising are developed as prototypes. Then, a new study to evaluate the performance of these three access controls is conducted. The results of this study show that users value positively these attributes and find them useful to define sharing policies. However, the results also indicate that users require assistance when employing these two new attributes during sharing policy definition. Finally, a qualitative analysis informs about what features users find important in an access control. Based on this, users value the granularity and understandability that tie strength offers and how tags can speed up the privacy configuration process. We hypothesize that the discrete performance results obtained by tie strength and tags can point to a lack of experience dealing with these attributes in an access control.

Regarding multiuser privacy management, this these proposes a novel model for representing and reasoning about multiuser scenarios, employing three types of features: contextual factors, user preferences, and user arguments. A crowdsourcing study is conducted to find important factors that potentially influence the inference of sharing decisions. Via a series of multinomial logistic regression models, this thesis shows that all three feature types influence the optimal sharing policy in a multiuser scenario. Additionally, the thesis analyzes what elements make scenarios harder to solve. The findings indicate that conflicts between arguments and preferences is the main reason why some scenarios have outcomes that are difficult to predict.

The results obtained in the previous study indicate that users interpret and prioritize arguments subjectively. This suggests that a tool that aims at helping users deal with multiuser conflicts must consider users' characteristics. Based on this idea, this thesis introduces Muppet, a sharing policy recommender for multiuser scenarios that aims at providing recommendations with low user effort. To achieve this goal, Muppet works incrementally and divides the recommendation process into three rounds. To create the recommendations, Muppet employs machine learning classifiers that exploit a variety of features: contextual factors, user preferences, user demographics, and group characteristics. After bootstrapping Muppet via crowdsourcing, an evaluation of its performance shows that Muppet outperforms baselines approaches. Moreover, the result also indicate that Muppet achieves acceptable accuracy scores when some contextual features are unknown. Therefore, after the bootstrapping process, Muppet is capable of providing acceptable recommendations with minimum user effort.

Bibliography

[1] Acquisti, A. and Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *Proceedings of the 6th International Conference on Privacy Enhancing Technologies (PET)*, pages 36–58, Cambridge, UK. Springer-Verlag.

[2] Amershi, S., Fogarty, J., and Weld, D. (2012). Regroup: Interactive machine learning for on-demand group creation in social networks. In *Proc. of the ACM Conference on Human Factors in Computing Systems*. ACM.

[3] Barnes, S. (2006). A privacy paradox: Social networking in the united states. *First Monday*, **11**(9), 11–15.

[4] Barth, A., Datta, A., Mitchell, J. C., and Nissenbaum, H. (2006). Privacy and contextual integrity: Framework and applications. In *Proceedings of the 2006 IEEE Symposium on Security and Privacy*, SP, pages 184–198, Oakland, CA. IEEE Computer Society.

[5] Bejugam, R. and LeFevre, K. (2011). enlist: Automatically simplifying privacy policies. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pages 620–627. IEEE.

[6] Besmer, A. and Lipford, H. (2010). Moving beyond untagging: Photo privacy in a tagged world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 1563–1572, Atlanta. ACM.

[7] Bilge, L., Strufe, T., Balzarotti, D., and Kirda, E. (2009). All your contacts are belong to us: automated identity theft attacks on social networks. In *Proceedings of the 18th international conference on World wide web*, pages 551–560. ACM.

[8] Bird, C., Gourley, A., Devanbu, P., Gertz, M., and Swaminathan, A. (2006). Mining email social networks. In *Proceedings of the 2006 International Workshop on Mining Software Repositories*, MSR '06, pages 137–143, New York, NY, USA. ACM.

[9] Bischoff, K. (2012). We love rock 'n' roll: Analyzing and predicting friendship links in last.fm. In *Proceedings of the 4th Annual ACM Web Science Conference*, WebSci '12, pages 47–56, New York, NY, USA. ACM.

[10] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, **2008**(10), P10008.

[11] Bonneau, J., Anderson, J., and Church, L. (2009). Privacy suites: Shared privacy for social networks. In *Symposium on Usable Privacy and Security (SOUPS)*. Citeseer.

[12] Boyd, D. and Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, **13**(1), 210–230.

[13] Boyd, D. and Hargittai, E. (2010). Facebook privacy settings: Who cares? *First Monday*, **15**(8).

[14] Boyd, D. and Heer, J. (2006). Profiles as conversation: Networked identity performance on friendster. In *System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on*, volume 3, pages 59c–59c. IEEE.

[15] Brand, A., Allen, L., Altman, M., Hlava, M., and Scott, J. (2015). Beyond authorship: attribution, contribution, collaboration, and credit. *Learned Publishing*, **28**(2), 151–155.

[16] Brandtzæg, P. B., Lüders, M., and Skjetne, J. H. (2010). Too many facebook "friends"? content sharing and sociability versus the need for privacy in social network sites. *International Journal of Human-Computer Interaction*, **26**(11-12), 1006–1030.

[17] Breiman, L. (2001). Random forests. *Machine Learning*, **45**(1), 5–32.

[18] Bruns, G., Fong, P., Siahaan, I., and Huth, M. (2012). Relationship-based access control: its expression and enforcement through hybrid logic. In *Proceedings of the second ACM conference on Data and Application Security and Privacy*, pages 117–124. ACM.

[19] Burt, R. (1995). *Structural holes: The social structure of competition*. Harvard Univ Pr.

[20] Carminati, B. and Ferrari, E. (2011). Collaborative access control in on-line social networks. In *Proceedings of the 7th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, pages 231–240.

[21] Carminati, B., Ferrari, E., and Perego, A. (2006). Rule-based access control for social networks. In *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops*, pages 1734–1744. Springer.

[22] Carminati, B., Ferrari, E., and Perego, A. (2009). Enforcing access control in web-based social networks. *ACM Transactions on Information and System Security (TISSEC)*, **13**(1), 6.

[23] Carminati, B., Ferrari, E., Heatherly, R., Kantarcioglu, M., and Thuraisingham, B. (2011). Semantic web-based social network access control. *computers*

& security, **30**(2-3), 108–115. Special Issue on Access Control Methods and Technologies.

[24] Cheek, G. P. and Shehab, M. (2012). Policy-by-example for online social networks. In *Proc. of the 17th ACM SACMAT*, SACMAT '12, pages 23–32, New York, NY, USA. ACM.

[25] Cheng, Y., Park, J., and Sandhu, R. (2012). A user-to-user relationship-based access control model for online social networks. In N. Cuppens-Boulahia, F. Cuppens, and J. Garcia-Alfaro, editors, *Data and Applications Security and Privacy XXVI*, volume 7371 of *Lecture Notes in Computer Science*, pages 8–24. Springer Berlin Heidelberg.

[26] Cooke, M. and Buckley, N. (2008). Web 2.0, social networks and the future of market research. *International Journal of Market Research*, **50**(2), 267–292.

[27] Cranor, L. and Garfinkel, S. (2005). *Security and Usability*. O'Reilly Media, Inc.

[28] Criado, N. and Such, J. M. (2015). Implicit contextual integrity in online social networks. *Information Sciences*, **325**, 48–69.

[29] Criado, N. and Such, J. M. (2016). Selective norm monitoring. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, pages 208–214.

[30] Culotta, A., Bekkerman, R., and Mccallum, A. (2004). Extracting social networks and contact information from email and the web. In *In Proceedings of CEAS-1*.

[31] Daniel, W. (1990). Kruskal-wallis one-way analysis of variance by ranks. *Applied Nonparametric Statistics,*, pages 226–230.

[32] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**(1), 1–38.

[33] Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, **10**(7), 1895–1923.

[34] Ding, L., Steil, D., Dixon, B., Parrish, A., and Brown, D. (2011). A relation context oriented approach to identify strong ties in social networks. *Knowledge-Based Systems*, **24**(8), 1187–1195.

[35] Dong, C., Jin, H., and Knijnenburg, B. (2015). Predicting privacy behavior on online social networks. In *International AAAI Conference on Web and Social Media*.

[36] Duck, S. (2007). *Human relationships*. Sage Publications Ltd.

[37] Dunbar, R. I. M. (2016). Do online social media cut through the constraints that limit the size of offline social networks? *Royal Society Open Science*, **3**(1).

[38] Ellison, N., Steinfield, C., and Lampe, C. (2007). The benefits of facebook friends: Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, **12**(4), 1143–1168.

[39] Fang, L. and LeFevre, K. (2010). Privacy wizards for social networking sites. In *Proceedings of the 19th international conference on World wide web*, pages 351–360, Raleigh, North Carolina. ACM, ACM.

[40] Felt, A. and Evans, D. (2008). Privacy protection for social networking apis. *2008 Web 2.0 Security and Privacy (W2SP'08)*.

[41] Fogues, R., Such, J., Espinosa, A., and Garcia-Fornes, A. (2014). Bff: A tool for eliciting tie strength and user communities in social networking services. *Information Systems Frontiers*, **16**(2), 225–237.

[42] Fogues, R. L., Such, J. M., Minguet, A. E., and García-Fornes, A. (2015). Open challenges in relationship-based privacy mechanisms for social network services. *International Journal of Human-Computer Interaction*, **31**(5), 350–370.

[43] Fogues, R. L., Such, J. M., Minguet, A. E., and García-Fornes, A. (2017a). Exploring the viability of tie strength and tags in access controls for photo sharing. In *Proceedings of the 32nd ACM Symposium on Applied Computing*, SAC '17. ACM.

[44] Fogues, R. L., Murukannaiah, P. K., Such, J. M., and Singh, M. (2017b). Sharing policies in multiuser privacy scenarios: Incorporating context, preferences, and arguments in decision making. *ACM Transactions on Computer-Human Interaction*.

[45] Fogues, R. L., Murukannaiah, P. K., Such, J. M., and Singh, M. (To be published 2017c). Sosharp: Recommending sharing policies in multiuser privacy scenarios. *IEEE Internet Computing*.

[46] Fong, P. (2011). Relationship-based access control: protection model and policy language. In *Proceedings of the first ACM conference on Data and application security and privacy*, pages 191–202. ACM.

[47] Fong, P., Anwar, M., and Zhao, Z. (2009). A privacy preservation model for facebook-style social network systems. In M. Backes and P. Ning, editors, *Computer Security – ESORICS 2009*, volume 5789 of *Lecture Notes in Computer Science*, pages 303–320. Springer Berlin Heidelberg.

[48] Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, **486**(3-5), 75–174.

[49] Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I. H., and Trigg, L. (2010). *Weka-A Machine Learning Workbench for Data Mining*, pages 1269–1277. Springer US, Boston, MA.

[50] Gadiraju, U., Kawase, R., Dietze, S., and Demartini, G. (2015). Understanding malicious behavior in crowdsourcing platforms: The case of online surveys. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI, pages 1631–1640, Seoul. ACM.

[51] Gates, C. (2007). Access control requirements for web 2.0 security and privacy. *IEEE Web*, **2**(0).

[52] Gilbert, E. (2012). Predicting tie strength in a new medium. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, pages 1047–1056, New York, NY, USA. ACM, ACM.

[53] Gilbert, E. and Karahalios, K. (2009). Predicting tie strength with social media. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 211–220, New York, NY, USA. ACM, ACM.

[54] Girvan, M. and Newman, M. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, **99**(12), 7821.

[55] Good, P. I. and Hardin, J. W. (2006). *Common Errors in Statistics (And How to Avoid Them)*. Wiley, New York.

[56] Granovetter, M. (1973). The strength of weak ties. *American journal of sociology*, **78**(6), 1.

[57] Greene, K., Derlega, V., and Mathews, A. (2006). Self-disclosure in personal relationships. *The Cambridge handbook of personal relationships*, **–**, 409–427.

[58] Gross, R. and Acquisti, A. (2005). Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80. ACM.

[59] Gujarati, D. N. and Porter, D. C. (2009). *Basic Econometrics*, chapter 10, pages 120–150. McGraw-Hill/Irwin, New York, Fifth edition. Multicollinearity: What happens if the regressors are correlated?

[60] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations Newsletter*, **11**(1), 10–18.

[61] Hart, M., Castille, C., Johnson, R., and Stent, A. (2009). Usable privacy controls for blogs. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, volume 4 of *CSE*, pages 401–408. IEEE, IEEE Computer Society.

[62] Hildebrandt, M. (2008). Defining profiling: A new type of knowledge? In M. Hildebrandt and S. Gutwirth, editors, *Profiling the European Citizen*, pages 17–45. Springer Netherlands.

[63] Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures*. John Wiley & Sons, New York.

[64] Hogan, B. (2010). The presentation of self in the age of social media: Distinguishing performances and exhibitions online. *Bulletin of Science, Technology and Society*, **30**(6), 377–386.

[65] Hogben, G. (2007). Security issues and recommendations for online social networks. *Position Paper ENISA European Network and Information Security Agency*, **80211**(1).

[66] Hollander, M. and Wolfe, D. A. (1999). *Nonparametric Statistical Methods*. Wiley, New York.

[67] Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**(2), 65–70.

[68] Houghton, D. J. and Joinson, A. N. (2010). Privacy, social network sites, and social relations. *Journal of Technology in Human Services*, **28**(1-2), 74–94.

[69] Hu, H. and Ahn, G. (2011). Multiparty authorization framework for data sharing in online social networks. *Data and Applications Security and Privacy XXV*, pages 29–43.

[70] Hu, H., Ahn, G.-J., and Jorgensen, J. (2011). Detecting and resolving privacy conflicts for collaborative data sharing in online social networks. In *Proceedings*

of the 27th Annual Computer Security Applications Conference (ACSAC), pages 103–112, Orlando. ACM.

[71] Hu, H., Ahn, G.-J., and Jorgensen, J. (2013). Multiparty access control for online social networks: Model and mechanisms. *IEEE Transactions on Knowledge and Data Engineering*, **25**(7), 1614–1627.

[72] Hu, H., Ahn, G.-J., Zhao, Z., and Yang, D. (2014). Game theoretic analysis of multiparty access control in online social networks. In *Proceedings of the 19th ACM Symposium on Access Control Models and Technologies (SACMAT)*, SACMAT '14, pages 93–102, London, Ontario. ACM.

[73] Ilia, P., Polakis, I., Athanasopoulos, E., Maggi, F., and Ioannidis, S. (2015). Face/Off: Preventing privacy leakage from photos in social networks. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pages 781–792, Denver, Colorado. ACM.

[74] Imlawi, J. and Gregg, D. (2014). Engagement in online social networks: The impact of self-disclosure and humor. *International Journal of Human-Computer Interaction*, **30**(2), 106–125.

[75] Javed, Y. and Shehab, M. (2013). Access control policy misconfiguration detection in online social networks. In *Social Computing (SocialCom), 2013 International Conference on*, pages 544–549.

[76] Johnson, M., Egelman, S., and Bellovin, S. M. (2012). Facebook and privacy: it's complicated. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, SOUPS '12, pages 9:1–9:15, New York, NY, USA. ACM, ACM.

[77] Jones, S. and O'Neill, E. (2010). Feasibility of structural network clustering for group-based privacy control in social networks. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, page 9. ACM.

[78] Joshi, P. and Kuo, C.-C. (2011). Security and privacy in online social networks: A survey. In *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, pages 1–6.

[79] Kagal, L., Hanson, C., and Weitzner, D. (2008). Using dependency tracking to provide explanations for policy management. In *Policies for Distributed Systems and Networks, 2008. POLICY 2008. IEEE Workshop on*, pages 54–61.

[80] Kahanda, I. and Neville, J. (2009). Using transactional information to predict link strength in online social networks. In *Proceedings of the Third International Conference on Weblogs and Social Media (ICWSM)*.

[81] Kairam, S., Brzozowski, M., Huffaker, D., and Chi, E. (2012). Talking in circles: Selective sharing in Google+. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, CHI '12, pages 1065–1074, Austin. ACM.

[82] Karjoth, G., Schunter, M., and Waidner, M. (2003). Platform for enterprise privacy practices: Privacy-enabled management of customer data. In R. Dingledine and P. Syverson, editors, *Privacy Enhancing Technologies*, volume 2482 of *Lecture Notes in Computer Science*, pages 69–84. Springer Berlin Heidelberg.

[83] Klemperer, P., Liang, Y., Mazurek, M., Sleeper, M., Ur, B., Bauer, L., Cranor, L. F., Gupta, N., and Reiter, M. (2012). Tag, you can see it!: using tags for access control in photo sharing. In *Proc. CHI*, pages 377–386, Austin.

[84] Kökciyan, N. and Yolum, P. (2016). Priguard: A semantic approach to detect privacy violations in online social networks. *IEEE Transactions on Knowledge and Data Engineering*, **28**(10), 2724–2737.

[85] Koroleva, K. and Bolufé Röhler, A. (2012). Reducing information overload: Design and evaluation of filtering & ranking algorithms for social networking

sites. In *Proceedings of the ECIS 2012 European Conference on Information Systems*.

[86] Kucuktunc, O., Sevil, S. G., Tosun, A. B., Zitouni, H., Duygulu, P., and Can, F. (2008). *Tag Suggestr: Automatic Photo Tag Expansion Using Visual Information for Photo Sharing Websites*, pages 61–73. Springer Berlin Heidelberg, Berlin, Heidelberg.

[87] Lampinen, A., Lehtinen, V., Lehmuskallio, A., and Tamminen, S. (2011). We're in it together: Interpersonal management of disclosure in social network services. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 3217–3226, Vancouver. ACM.

[88] Lancichinetti, A. and Fortunato, S. (2009). Community detection algorithms: A comparative analysis. *Phys. Rev. E*, **80**, 056117.

[89] Lancichinetti, A., Fortunato, S., and Kertész, J. (2009). Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics*, **11**(3), 033015.

[90] Li, Q., Li, J., Wang, H., and Ginjala, A. (2011). Semantics-enhanced privacy recommendation for social networking sites. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on*, pages 226–233. IEEE.

[91] Lilliefors, H. W. (1967). On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, **62**(318), 399–402.

[92] Lilliefors, H. W. (1969). On the kolmogorov-smirnov test for the exponential distribution with mean unknown. *Journal of the American Statistical Association*, **64**(325), 387–389.

[93] Lin, N., Ensel, W., and Vaughn, J. (1981). Social resources and strength of ties: Structural factors in occupational status attainment. *American sociological review*, pages 393–405.

[94] Lipford, H., Besmer, A., and Watson, J. (2008). Understanding privacy settings in facebook with an audience view. In *Proceedings of the 1st Conference on Usability, Psychology, and Security*, pages 1–8. USENIX Association Berkeley, CA, USA.

[95] Lipford, H., Watson, J., Whitney, M., Froiland, K., and Reeder, R. (2010). Visual vs. compact: A comparison of privacy policy interfaces. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 1111–1114. ACM.

[96] Lippi, M. and Torroni, P. (2016). Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology*, **16**(2), 10:1–10:25.

[97] Liu, G., Wang, Y., and Orgun, M. (2010). Optimal social trust path selection in complex social networks. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence, AAAI*, pages 1391–1398.

[98] Liu, K. and Terzi, E. (2009). A framework for computing the privacy scores of users in online social networks. In *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on*, pages 288–297. IEEE.

[99] Mahmood, S. and Desmedt, Y. (2011). Poster: preliminary analysis of google+'s privacy. In *Proceedings of the 18th ACM conference on Computer and communications security*, pages 809–812. ACM.

[100] Mancini, C., Rogers, Y., Bandara, A. K., Coe, T., Jedrzejczyk, L., Joinson, A. N., Price, B. A., Thomas, K., and Nuseibeh, B. (2010). ContraVision: Exploring users' reactions to futuristic technology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 153–162. ACM.

[101] Marwick, A. and Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, **13**(1), 114–133.

[102] Matsuo, Y., Mori, J., Hamasaki, M., Nishimura, T., Takeda, H., Hasida, K., and Ishizuka, M. (2007). Polyphonet: an advanced social network extraction system from the web. *Web Semantics: Science, Services and Agents on the World Wide Web*, **5**(4), 262–278. World Wide Web Conference 2006Semantic Web Track.

[103] Mazzia, A., LeFevre, K., and Adar, E. (2011). The pviz comprehension tool for social network privacy settings. *University of Michigan CSE Technical Report CSE-TR-570-11*.

[104] Munemasa, T. and Iwaihara, M. (2011). Trend analysis and recommendation of users' privacy settings on social networking services. *Social Informatics*, pages 184–197.

[105] Murukannaiah, P. and Singh, M. (2012). Platys social: Relating shared places and private social circles. *Internet Computing, IEEE*, **16**(3), 53–59.

[106] Murukannaiah, P. K. and Singh, M. P. (2015). Platys: An active learning framework for place-aware application development and its evaluation. *ACM Transactions on Software Engineering and Methodology*, **24**(3), 1–33.

[107] Murukannaiah, P. K., Kalia, A. K., Telang, P. R., and Singh, M. P. (2015). Resolving goal conflicts via argumentation-based analysis of competing hypotheses. In *Proceedings of the 23rd IEEE International Requirements Engineering Conference*, pages 156–165, Ottawa.

[108] Murukannaiah, P. K., Ajmeri, N., and Singh, M. P. (2016). Engineering privacy in social applications. *IEEE Internet Computing*, **20**(2), 72–76.

[109] Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, **79**, 119.

[110] Paolacci, G., Chandler, J., and Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, **5**(5), 411–419.

[111] Paradesi, S., Liccardi, I., Kagal, L., and Pato, J. (2013). A semantic framework for content-based access controls. In *Social Computing (SocialCom), 2013 International Conference on*, pages 624–629.

[112] Pearson, S. and Mont, M. (2011). Sticky policies: An approach for managing privacy across multiple parties. *Computer*, **44**(9), 60–68.

[113] Peddinti, S. T., Korolova, A., Bursztein, E., and Sampemane, G. (2014). Cloak and swagger: Understanding data sensitivity through the lens of user anonymity. In *2014 IEEE Symposium on Security and Privacy*, pages 493–508. IEEE.

[114] Peer, E., Vosgerau, J., and Acquisti, A. (2014). Reputation as a sufficient condition for data quality on amazon mechanical turk. *Behavior Research Methods*, **46**(4), 1023–1031.

[115] Pergament, D., Aghasaryan, A., Ganascia, J., and Betgé-Brezetz, S. (2011). Forps: friends-oriented reputation privacy score. In *Proceedings of the First International Workshop on Security and Privacy Preserving in e-Societies*, pages 19–25. ACM.

[116] Pike, G. (2011). Fired over facebook. *Information Today*, **28**(4), 26–26.

[117] Quercia, D., Lambiotte, R., Kosinski, M., Stillwell, D., and Crowcroft, J. (2012). The personality of popular facebook users. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW'12)*.

[118] Quinlan, J. R. (1993). *C4.5: programs for machine learning*, volume 1. Morgan kaufmann.

[119] Rana, J., Kristiansson, J., and Synnes, K. (2010). Enriching and simplifying communication by social prioritization. In *Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on*, pages 336–340. IEEE.

[120] Reeder, R., Bauer, L., Cranor, L., Reiter, M., Bacon, K., How, K., and Strong, H. (2008). Expandable grids for visualizing and authoring computer security policies. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 1473–1482. ACM.

[121] Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science*, **6**(1), 15–32.

[122] Rosen, D. and Chu, K. (2011). The utility of communication network ties: Reconceptualizing the social network tie measure. In *System Sciences (HICSS), 2011 44th Hawaii International Conference on*, pages 1–8. IEEE.

[123] Rosvall, M. and Bergstrom, C. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, **105**(4), 1118–1123.

[124] Samarati, P. and Sweeney, L. (1998). Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report, Technical report, SRI International.

[125] Shehab, M., Cheek, G., Touati, H., Squicciarini, A., and Cheng, P. (2010). Learning based access control in online social networks. In *Proceedings of the 19th international conference on World wide web*, pages 1179–1180. ACM.

[126] Shen, K., Song, L., Yang, X., and Zhang, W. (2010). A hierarchical diffusion algorithm for community detection in social networks. In *Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2010 International Conference on*, pages 276–283. IEEE.

[127] Shih, D.-H., Hsu, S.-F., Yen, D. C., and Lin, C.-C. (2012). Exploring the individual's behavior on self-disclosure online. *International Journal of Human-Computer Interaction*, **28**(10), 627–645.

[128] Shvartzshnaider, Y., Tong, S., Wies, T., Kift, P., Nissenbaum, H., Subramanian, L., and Mittal, P. (2016). Learning privacy expectations by

crowdsourcing contextual informational norms. In *Proceedings of the Fourth AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*.

[129] Sierra, C. and Debenham, J. (2007). The LOGIC negotiation model. In *AAMAS '07: Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, pages 1–8. ACM.

[130] Siersdorfer, S., Minack, E., Deng, F., and Hare, J. S. (2010). Analyzing and predicting sentiment of images on the social web. In *Proceedings of the 18th ACM International Conference on Multimedia*, MM, pages 715–718, Firenze, Italy. ACM.

[131] Singh, M. P. (2013). Norms as a basis for governing sociotechnical systems. *ACM Transactions on Intelligent Systems and Technology*, **5**(1), 21:1–21:23.

[132] Sleeper, M., Balebako, R., Das, S., McConahy, A. L., Wiese, J., and Cranor, L. F. (2013). The post that wasn't: Exploring self-censorship on Facebook. In *Proceedings of the conference on Computer supported cooperative work (CSCW)*, pages 793–802. ACM.

[133] Squicciarini, A., Shehab, M., and Paci, F. (2009). Collective privacy management in social networks. In *Proceedings of the 18th international conference on World wide web*, pages 521–530, Madrid. ACM, ACM.

[134] Squicciarini, A., Sundareswaran, S., Lin, D., and Wede, J. (2011). A3p: adaptive policy prediction for shared images over popular content sharing sites. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*, HT, pages 261–270, New York. ACM, ACM.

[135] Squicciarini, A., Paci, F., and Sundareswaran, S. (2014). Prima: a comprehensive approach to privacy protection in social network sites. *annals of telecommunications - annales des télécommunications*, **69**(1-2), 21–36.

[136] Staddon, J., Huffaker, D., Brown, L., and Sedley, A. (2012). Are privacy concerns a turn-off? engagement and privacy in social networks. In *Symposium on Usable Privacy and Security (SOUPS)*, page 10. ACM.

[137] Steurer, M. and Trattner, C. (2013). Acquaintance or partner?: Predicting partnership in online and location-based social networks. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '13, pages 372–379, New York, NY, USA. ACM.

[138] Stieglitz, S. and Dang-Xuan, L. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, **29**(4), 217–248.

[139] Strater, K. and Lipford, H. (2008). Strategies and struggles with privacy in an online social networking community. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 1*, BCS-HCI '08, pages 111–119, Swinton, UK, UK. British Computer Society, British Computer Society.

[140] Stutzman, F. and Kramer-Duffield, J. (2010). Friends only: examining a privacy-enhancing behavior in facebook. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI, pages 1553–1562, New York. ACM, ACM.

[141] Stutzman, F., Vitak, J., Ellison, N., Gray, R., and Lampe, C. (2012). Privacy in interaction: Exploring disclosure and social capital in facebook. In *International AAAI Conference on Web and Social Media*.

[142] Such, J. M. and Criado, N. (2014). Adaptive conflict resolution mechanism for multi-party privacy management in social media. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*, pages 69–72. ACM.

[143] Such, J. M. and Criado, N. (2016). Resolving multi-party privacy conflicts

in social media. *IEEE Transactions on Knowledge and Data Engineering*, **28**(7), 1851–1863. To appear.

[144] Such, J. M. and Rovatsos, M. (2016). Privacy policy negotiation in social media. *ACM Transactions on Autonomous and Adaptive Systems*, **11**(4), 1–29.

[145] Such, J. M., Espinosa, A., García-Fornes, A., and Sierra, C. (2012). Self-disclosure decision making based on intimacy and privacy. *Information Sciences*, **211**(0), 93 – 111.

[146] Such, J. M., Espinosa, A., and García-Fornes, A. (2014). A survey of privacy in multi-agent systems. *Knowledge Engineering Review*, **29**(03), 314–344.

[147] Thomas, K., Grier, C., and Nicol, D. (2010). unfriendly: Multi-party privacy risks in social networks. In *Privacy Enhancing Technologies*, pages 236–252, Berlin. Springer, Springer-Verlag.

[148] Toch, E., Wang, Y., and Cranor, L. F. (2012). Personalization and privacy: A survey of privacy risks and remedies in personalization-based systems. *User Modeling and User-Adapted Interaction*, **22**(1-2), 203–220.

[149] Walton, D., Reed, C., and Macagno, F. (2008). *Argumentation Schemes*. Cambridge University Press.

[150] Wang, Y., Norcie, G., Komanduri, S., Acquisti, A., Leon, P. G., and Cranor, L. F. (2011). I regretted the minute I pressed share: A qualitative study of regrets on Facebook. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, page 10. ACM.

[151] Wang, Y., Leon, P. G., Scott, K., Chen, X., Acquisti, A., and Cranor, L. F. (2013). Privacy nudges for social media: An exploratory facebook study. In *Proceedings of the 22Nd International Conference on World Wide Web Companion*, WWW '13 Companion, pages 763–770, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

[152] Warren, S. D. and Brandeis, L. D. (1890). The right to privacy. *Harvard Law Review*, **4**(5), pp. 193–220.

[153] Watson, J., Lipford, H., and Besmer, A. (2015). Mapping user preference to privacy default settings. *ACM Transactions on Computer-Human Interaction*, **22**(6), 32:1–32:20.

[154] Wellman, B. and Wortley, S. (1990). Different strokes from different folks: Community ties and social support. *American journal of Sociology*, pages 558–588.

[155] Westin, A. (1968). Privacy and freedom. *Washington and Lee Law Review*, **25**(1), 166.

[156] Wiese, J., Kelley, P., Cranor, L., Dabbish, L., Hong, J., and Zimmerman, J. (2011). Are you close with me? are you nearby? investigating social groups, closeness, and willingness to share. In *Proceedings of the 13th international conference on Ubiquitous computing*, UbiComp, pages 197–206, New York. ACM, ACM.

[157] Williams, M. and Williamson, J. (2006). Combining argumentation and Bayesian nets for breast cancer prognosis. *Journal of Logic, Language, and Information*, **15**(1–2), 155–178.

[158] Wishart, R., Corapi, D., Marinovic, S., and Sloman, M. (2010). Collaborative privacy policy authoring in a social networking context. In *Policies for Distributed Systems and Networks (POLICY), 2010 IEEE International Symposium on*, pages 1–8. IEEE.

[159] Wisniewski, P., Lipford, H., and Wilson, D. (2012). Fighting for my space: Coping mechanisms for SNS boundary regulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 609–618, Austin. ACM.

[160] Wu, A., DiMicco, J., and Millen, D. (2010). Detecting professional versus personal closeness using an enterprise social network site. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 1955–1964. ACM.

[161] Xiang, R., Neville, J., and Rogati, M. (2010). Modeling relationship strength in online social networks. In *Proceedings of the 19th international conference on World wide web*, pages 981–990. ACM.

[162] Yao, M., Rice, R., and Wallis, K. (2007). Predicting user concerns about online privacy. *Journal of the American Society for Information Science and Technology*, **58**(5), 710–722.

[163] Yeung, C., Kagal, L., Gibbins, N., and Shadbolt, N. (2009). Providing access control to online photo albums based on tags and linked data. In *Proceedings of the AAAI Spring Symposium on Social Semantic Web: Where Web*, volume 2.

[164] Yildiz, H. and Kruegel, C. (2012). Detecting social cliques for automated privacy control in online social networks. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, pages 353 –359.

[165] You, Q., Luo, J., Jin, H., and Yang, J. (2015). Robust image sentiment analysis using progressively trained and domain transferred deep networks. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI, pages 381–388. AAAI Press.

[166] Zhang, C., Sun, J., Zhu, X., and Fang, Y. (2010). Privacy and security for online social networks: challenges and opportunities. *Network, IEEE*, **24**(4), 13–18.

[167] Zheleva, E. and Getoor, L. (2011). Privacy in social networks: A survey. In C. C. Aggarwal, editor, *Social Network Data Analytics*, pages 277–306. Springer US.