A Time-Evolution Model for the Privacy Degree of Information Disseminated in Online Social Networks

Lotfi ben Othmane* and Harold Weffers

Dept of Mathematics and Computer Science,
Eindhoven University of Technology,
Eindhoven, Netherlands
Email:l.ben.othmane@tue.nl
Email:h.t.g.weffers@tue.nl
*Corresponding author

Pelin Angin and Bharat Bhargava

Dept of Computer Science,
Purdue University,
West Lafayette, IN, USA
Email: pangin@cs.purdue.edu
Email: bb@cs.purdue.edu

Abstract: People tend to share private information with their friends on Online Social Networks (OSNs). The common position is that the shared information eventually reaches all users of the network since OSNs exhibit the small-world property. However, dissemination of private information in an OSN exhibits a set of factors that need to be accounted for in order to create more realistic models of the evolution of the privacy degree of information disseminated in an OSN. Among these factors are relationship strength between communicating users, influence of neighbors (i.e., friends), users' adoption of new information, change of information, and dynamics of the structure of OSNs.

This paper proposes a time series model for measuring the privacy of information disseminated in an OSN using the factors listed above. It shows through simulating the dissemination of private information in an OSN that the privacy of information does not vanish, but in most cases declines to a saturation level related to the information dissemination factors. The results also show how likely a user can get the information when the factors are accounted for.

Keywords: Online Social Network, Privacy, Data dissemination

Reference to this paper should be made as follows: Lotfi ben Othmane, Harold Weffers, Pelin Angin and Bharat Bhargava (2013) 'A Time-Evolution Model for the Privacy Degree of Information

Disseminated in Online Social Networks', *Int. J. Communication Networks and Distributed Systems*, Vol. x, No. x, pp.xxx-xxx.

Biographical notes: Lotfi ben Othmane received his Ph.D. degree from Western Michigan University (WMU), USA, in 2010, M.S. degree in Computer Science from University of Sherbrooke, Canada, in 2000, and B.S degree from University of Sfax, Tunisia, in 1995. He has extensive experience in the industry as Programmer, Software Architect, System Analyst and Technology Manager in Tunisia, Canada and USA. He is currently a Postdoc at the Laboratory for Quality Software (LaQuSo), Eindhoven University of Technology (TU/e), Netherlands. Previously, he was a Visiting Assistant Professor at Kalamazoo College, USA. Dr. ben Othmane′s main research topics are: cross domain information sharing, development of secure systems using Agile approach, and safety and security in connected vehicles.

Harold Weffers is the Director of the Laboratory for Quality Software (LaQuSo) at Eindhoven University of Technology. He received his M.Sc. degree in Computer Science in 1993 and his PDEng degree in Software Technology in 1995. After working for the Royal Netherlands Navy and Philips he joined the Eindhoven University of Technology in 1998 as Director of the Professional Doctorate in Engineering degree program in Software Technology. In 2008 he moved to his current position. He is currently also a member of the following committees: board of the School for Medical Physics and Engineering, Eindhoven, the advisory board of the PDEng Qualified Medical Informatics programme of the School for Medical Physics and Engineering Eindhoven, the core team of the PDEng Smart Energy Buildings and Cities programme (EIT KIC InnoEnergy), and the NEN Norm Committee on systems and software engineering (related to ISO/IEC JTC1-SC7). He authored and co-authored several publications, especially on software quality.

Pelin Angin is a Ph.D. Student at the Department of Computer Science at Purdue University, USA. She received her BS degree in Computer Engineering at Bilkent University, Turkey in 2007. Her research interests lie in the fields of Mobile-Cloud Computing, Cloud Computing Privacy and Data Mining. She is currently working under the supervision of Professor Bharat Bhargava on leveraging mobile-cloud computing for real-time context-awareness and development of algorithms to address the associated privacy issues.

Bharat Bhargava received the B.E. degree from the Indian Institute of Science and the M.S. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN. He is currently a Professor of computer science at Purdue University. His research involves mobile wireless networks, secure routing and dealing with malicious hosts, providing security in Service Oriented Architectures (SOA), adapting to attacks, and experimental studies. He is a Fellow of the IEEE Computer Society. His name has been included in the Book of Great Teachers at Purdue University. Moreover, he was selected by the student chapter of ACM at Purdue University for the Best Teacher Award.

1 Introduction

Online Social Network (OSN) sites such as Facebook (Facebook, 2012), Pinterest (Pinterest, 2012), and YouTube (YouTube, 2012) are web applications for maintaining social relationships, and sharing messages and content, which are in the form of text, image, video, web links, etc. The main components of an OSN are (cf. Mislove et al., 2007):

- *Users*: Individuals could register with an OSN and provide information about themselves (e.g., birthday, place of residence, phone number, and email address) which compose their profile; they become users of the OSN.
- *Groups*: Users can create and join special interest groups. They can post messages or upload content to the publishing space of the group.
- *Content*: Users of OSNs share content. Each user gets a publishing space (e.g., wall in Facebook), when he/she registers with the OSN, where he/she shares content and messages.
- *Links*: Users are connected using links. Links express online relationship, e.g., friendship, business contact, and common interests. In an OSN, the neighbors of a user are the set of users with whom he/she shares links.

Users of OSNs disclose information in the form of messages and content including their own private information: information that should not be public, such as, Social Security Number (SSN), birthday, pictures, address, credit card numbers, record of bank transactions, record of flights, preferred movies, and income tax reports with their neighbors. For instance, a Facebook user is able to share his/her private information by posting an image, a video, or a message on his/her wall. (We use the term *subject* to refer to the user the private information is about.) He/She could specify whether the content is visible to only his/her neighbors, or is public information–visible to any user who accesses the wall of the publisher. In this paper we consider private information shared by users with their friends; we do not consider private information disseminated to the public.

Private information disclosures in OSNs have privacy implications (Gross and Acquisti, 2005; Acquisti, 2004). Privacy is the right of an entity to be able to control when, how, to what extent, and for what purpose information about himself/herself is shared with others (cf. Westin, 1967), and to determine the degree to which the entity will interact with its environment (cf. (Shirey, 2007)). Privacy degree of information quantifies, in a scale from 0 to 1, how private the information is (more details are provided in Subsection 2.1).

Disclosed private information could be used, for example, in espionage: attackers obtain private and sensitive data about their opponents. They collect data from entities that obtained the information from the subject or through a dissemination network. Private information could be used, for example, to learn the business activities of a user. Disclosed private information could also be used for identity theft: An attacker may obtain a set of information about an individual which are sufficient to steal his identity. Identity thefts use identity information (information that help to uniquely identify a user, which includes name, birthday, place of residence, and SSN) to obtain, for example, financial advantages (e.g., use credit cards of the subject to buy goods and services). Moreover, the news reported cases where information disclosed in OSNs for specific context was used by specific users of OSNs–who received it–in other contexts, which affects the life of the subject of the information as in (Rosen, 2010).

OSNs exhibit the small-world property (Mislove et al., 2007). Small-world property refers to a short chain of nodes connecting even the most distant users of the network. A small-world phenomenon is known due to Milgram's finding (Milgram, 1967) that the average path length between two Americans is 6 hops. The small-world property of OSNs led to the common position–and belief–that OSNs cause the vanishing of the privacy of their users. A user of an OSN who discloses his/her private information to his neighbors loses his privacy: his friends will disseminate his private information and eventually all the network users will *know* the information. The belief is based on the following assumptions:

- users who receive private information from their neighbors believe it.
- users who believe private information about a user they receive from their neighbors disseminate it to their friends.

This belief contradicts the practice of the users of OSNs who tend to be selective in the information they disclose to their friends. For instance, Govani and Pashley (Govani and Pashley, 2012) show through a survey that students who use Facebook at Carnegie Mellon University (CMU), United States, are aware of the consequences of providing personally identifiable information (e.g., SSN, name, address, telephone, birthday) in OSNs but feel comfortable providing it. However, the users tend to not provide their mailbox, current address, and mobile and home phone numbers. There is a common position that users maintain some balance between privacy loss and benefits, such as gaining trust of friends when they disclose their private information. We believe such behavior has–beside this position–another cause: the students, unconsciously, believe that their privacy does not vanish when they disclose such information.

During the Tunisian revolution of 2011, activists used Facebook and twitter to share information. For example, they report about events and negotiations in the presidential palace and military bases, and activities of defected political and military leaders. The officers of the Tunisian intelligence agency joined the OSNs as friends to identify the activists and spy about their social activities–they send friendship invitations to potential activists. The officers failed to identify the (main) activists. Although, this paper does not analyze the causes of the failure, we believe it contributes to showing that the spying approach is not efficient.

The main problem that we investigate in this paper is how the privacy degree of private information disclosed by a subject to his/her neighbors changes over time considering several factors that affect information dissemination in OSNs.

This paper presents a model for the evolution of the privacy degree, over time, of a user of an OSN who discloses private information to his/her neighbors. Each user of the OSN who receives the information from one of his/her neighbors collects the perceptions of his/her neighbors (perception is a subjective opinion or belief about something (*Cambridge Advanced Learner's Dictionary.* 2012) by a specific user) and computes the influence of the perception of his/her neighbors. Then, the user combines his/her knowledge about the private information and the influence of the perception of the neighbors to get a new perception about the private information.

The model considers the effect of a set of factors on the privacy degree of the disseminated private information, which are: relationship strength, influence of neighbors, adoption of new information, change of information, and dynamics of the structure of OSNs. (Subsection 3.1 describes the factors.) The simulation extends the model by applying it on OSN graphs (OSN graphs are described in Subsection 2.2).

The model uses the factors representing facts related to the dissemination of private information in OSNs–e.g., information change (We provide examples of experimenting with the factors in Subsection 3.1), and assumes that these factors affect the privacy degree of disseminated private information to measure the evolution, over time, of the privacy degree of disseminated private information. It does not use datasets of logs of an OSN's use to measure the effects of the factors on the evolution of the privacy degree of disseminated information in OSNs. The paper measures the effects of the factors (i.e., what happens when we consider them) on this evolution and does not quantify the contribution of each of them. An analysis of such datasets would help to model the functions representing these factors and the contribution of each in the evolution of the privacy degree of disseminated private information in OSNs.

The main contributions of the paper are: (1) We show through simulating the dissemination of private information in an OSN that the privacy of the information does not vanish, but in most cases declines to a saturation point specific to values representing the information dissemination factors. (2) We show how likely a user of an OSN can get the private information disseminated in the OSN.

The paper is organized as follows. Section 2 provides a brief introduction of privacy, social network graphs, and information dissemination in OSNs. Section 3 describes the model that we use to describe the evolution of the privacy of a user who discloses his/her private information in an OSN. Section 4 describes the simulation of the proposed model and provides an analysis of the simulation results. Section 5 describes related work, and Section 6 concludes the paper with a discussion of future work.

2 Preliminaries

This section provides background about privacy, social network graphs, and information dissemination.

2.1 Privacy

We define *privacy degree* as follows: Given a piece of private information about a subject a of an OSN, the privacy degree of the information is the probability that a randomly selected user of the OSN does not know or does not believe the information.

In this work, we use privacy degree as a measure of the privacy of a piece of information in the scale of real numbers from 0 to 1. Privacy degree of a piece of private information is different from the privacy degree of a user who has a set of private information. The privacy degree of a user considers the privacy degrees of all his/her private information he/she disclosed in OSNs.

2.2 Social network graphs

Social networks are commonly modeled as directed graphs, where nodes correspond to the entities (or people) of the network and edges correspond to the relationships between the nodes. (For simplicity, we do not consider hypergraphs, where an edge can connect more than two nodes.) That is, a node has a directed edge to another node in the network if it can share information with it. The relationships represented by the edges can be of various types, such as “friend”, “colleague”, etc.

Real-world social networks exhibit (besides the small-world phenomenon) the non-random connection behavior, where there are well-defined locales with a high connection probability and the probability of connection between two vertices chosen at random is very low (Newman and Watts, 1999). Watts and Strogatz showed that real social networks possess random shortcuts that can link distant nodes (Watts and Strogatz, 1998). This finding suggests that the dissemination of information in a real-world social network is fast with information reaching even distant nodes faster than expected.

2.3 Information Dissemination in Social Networks

Dissemination (or diffusion) in social networks has been studied by many researchers on a variety of topics including spread of diseases (Newman, 2002) and spread of influence (Kempe, Kleinberg, and Tardos, 2003) among others. Below are two basic models commonly used in simulating information dissemination in OSNs (Mertsalov, Magdon-Ismail, and Goldberg, 2009). These models use a “disease spread” terminology, where being infected for a node means that the node has gained knowledge of the piece of information being disseminated in the network. The models are:

1. Linear threshold model: This model assigns a susceptibility threshold to each node and an influence threshold to each node pair at initialization. At each iteration of the algorithm, a node becomes infected if the sum of the influence thresholds of the pairs this node forms with other nodes exceeds the susceptibility threshold of the node.

2. Independent cascade model: This model assigns an infection probability to each node pair. At each iteration of the algorithm, a contagious (infected) node passes the infection to its neighbor with the infection probability assigned to the edge between the two nodes.

In this work, we use an information dissemination model that combines both models.

3 Description of the proposed model for measuring privacy degree of disseminated information in an OSN

This section describes the factors affecting the privacy degree of disseminated private information and the proposed model for measuring the privacy degree considering these factors.

3.1 Factors affecting the privacy degree of disseminated private information

The model considers the following factors related to the dissemination of private information in OSNs: relationship strength, influence of neighbors, adoption of new information, change of information, and dynamics of the structure of OSNs. The description of the factors follows.

Relationship strength. A user who receives private information about a subject propagates it to users with whom he/she has strong relationships. For instance, users of an OSN who have frequent and regular communications are friends who have a strong relationship. Users may not disseminate the private information of their friends to friends they have weak relationships with. The relationship strength factor enforces selection of a set of friends to receive the disclosed private information out of all the friends of the user.

We model the relationship strength of a link connecting 2 nodes, e.g., node j and node k, using a selection function. We use variable P^d_{jk} to model the strength of the relationship from node j to node k. The variable takes values in the range of [0, 1]. (The relationship strength of node k to node j is different from the relationship strength of node j to node k.) Then, we use variable D_{jk}, which represents the relationship strength threshold, to enforce selection of only strong relationships. That is, node j disseminates the private information $VPIF$ to node k only if $P^d_{jk} > D_{jk}$.

Influence of neighbors. Users tend to have different influences on their neighbors. Each link of an OSN connects two users who tend to influence each other. For instance, members of the family of a user may have more influence than a business customer; i.e. an OSN user is more likely to believe a piece of information when he/she receives it from a family member than to believe the same information when he/she receives it from a business customer.

We model the influence of node j on node k using variable I_{jk}, which has values in the range [0,1]. Variable I_{jk} follows a statistical distribution; however, the model is independent of the distribution.

Adoption of new information. Users of an OSN have different attitudes towards information they receive from their friends. For instance, users of OSNs have different behaviors toward rumors. A user who believes a piece of private information about a subject and receives new information from his/her neighbors about the same subject, may adopt the new private information, or may not adopt it–and continue to have the same perception about the information he/she has.

We model the adoption of new information by node k using weights in the range [0,1]. We use variable W_k to model the weights of the adoption of new information, which follows a statistical distribution. Note that the model is independent of the distribution.

Change of information. Private information about a user could change. The change could be natural or artificial. A user could change his/her residence location or job, which is a natural change of private information. A user could change his/her email address, which is an artificial change of private information. In both cases, the change makes previously disclosed information obsolete.

We model the change of information at any time step using a probability that models chances of change of information $VPIF$. We use the variable Pr^c to represent this probability. For instance, Pr^c equal to 0.1 indicates that the information changes at time steps multiples of 10–i.e., 10, 20, etc. Pr^c follows a statistical distribution; however, the model is independent of the distribution.

Dynamics of the structure of OSNs. The structures of OSNs change over time: new users are added, current users establish new relationships or lose some of their existing relationships, and some of current users leave the networks. For instance, a user of an OSN could change his/her residence location, school, or job; he/she establishes new friendships and loses some of his/her friendships. OSNs are dynamic since their structures (users and relationship between users) change over time.

To keep the model simple, we do not include this factor in the model. However, we use this factor in the simulation which extends the model.

3.2 Model for measuring privacy degree of disseminated information

In this section, we describe our proposed model for the time-evolution of the privacy degree of a piece of private information disclosed in an OSN. Note that we model OSNs using graphs, where nodes represent users and edges represent relationships between the users. Table 1 lists and describes the symbols used in the model.

Let OSN be a social network, where a user represented by node a discloses his/her piece of private information, $VPIF$, at time 0. At any time, each user of the network has one of the following 4 perception values of $VPIF$:

- *uk* (unknown): The user receives $VPIF$ from unknown source–not his/her neighbors, or does not know the information. (In this model, users receive information only from their neighbors, i.e. friends.)
- *bl* (believe): The user receives $VPIF$ from one or more of his/her neighbors and believes it is true: $VPIF_t$.
- *db* (disbelieve): The user receives $VPIF$ from his/her neighbors; however, he/she believes it is wrong: $VPIF_f$.
- *uc* (uncertain): The user receives $VPIF$ from his/her neighbors; however, he/she is undecided whether to believe it or not.

Note that a user of the network who does not receive $VPIF$ cannot disseminate his/her perception about it. A user who receives $VPIF$ *may* disseminate his/her perception to a set of his/her neighbors.

Table 1 List of symbols used in the model.

D_{jk}	Relationship strength threshold.
I_{jk}	Influence of node j on node k.
N	Number of nodes of the network.
N_k	Set of neighbors of node k.
$P(t)$	Probability that any user of the OSN knows the correct value of $VPIF$ at time t.
Pr^c	Probability for the change of $VPIF$.
P_{jk}^d	Relationship strength from node j to node k.
S_{max}	The minimum perception degree required for a node to believe $VPIF$ and set its perception to bl.
S_{min}	The maximum perception degree required for a node to choose to disbelieve $VPIF$ and set its perception to db.
$S_k(t)$	Perception degree (real-valued) of $VPIF$ by node k at time t considering the possibility of information change.
$S_k^+(t)$	Perception degree (real-valued) of $VPIF$ by node k at time t considering the adoption weight of node k.
$S_k^*(t)$	Perception (real-valued) of influence of neighbors of node k at time t regarding information $VPIF$.
$S_k^c(t)$	Perception value (categorical) of $VPIF$ by node k at time t considering a change of $VPIF$ by its subject.
$S_k^d(t)$	Perception value (categorical) of $VPIF$ by node k at time t.
U	Set of all users of the network.
$VPIF$	Private piece of information about node a. We use the following notations to describe the validity of the information: $VPIF_i$ for the initial true value, $VPIF_f$ for false value, $VPIF_t$ for true value, and $VPIF_c$ for true changed (updated) value.
W_k	Weight of the adoption of new information.

Equation 1, below, shows that, at time 0, each node of the OSN has perception uk of $VPIF$, except node a–the subject of the information, which has perception bl.

$$S_k^d(0) = \begin{cases} bl \text{ if } k = a \\ uk \text{ if } k! = a \end{cases} \tag{1}$$

Effect of strength of relationships and influence of neighbors. At time t, node k has perception $S_k^d(t)$ about $VPIF$. The perception of node k regarding $VPIF$ at time $t+1$ combines the perception degrees of the neighbors of node k that have strong relationships with it, while considering the influence of each neighbor. Equation 2 formulates the perception degree of node k about $VPIF$, at time $t+1$. This perception degree is the sum of the perception degrees of the neighbors of node k, which share strong relationships with k, weighted by the influence of each neighbor. The result is normalized to output a value in the range [-1,1], which could be mapped to one of the 4 perception values for $VPIF$.

$$S_k^*(t+1) = \frac{\sum_{j \in N_k} [(I_{jk} \times S_j^d(t))/P_{jk}^d > D_{jk}]}{|j \in N_k/P_{jk}^d > D_{jk}|} \tag{2}$$

Adoption of new information. Equation 3 below computes $S_k^+(t+1)$, perception degree of node k about $VPIF$ at time $t+1$, considering the adoption of new information factor. It combines the perception degree of node k at time step t and the sum of the perception degrees of the neighbors of the node at time step $t+1$. The aggregation function, a mean weighted function, uses the adoption of new information weight, W_k. In extreme cases, a user of the network can adopt the perception degrees of his/her neighbors and ignore his/her own perception degree about $VPIF$, in which case it sets the weight to 1.

$$S_k^+(t+1) = (W_k \times S_k^*(t+1)) + ((1 - W_k) \times S_k^d(t)) \tag{3}$$

$S_k^+(.)$ is the perception degree of node k about $VPIF$ considering relationship strength, node influence and adoption of new information factors. Equation 4 formulates–using a membership function–the decision on the perception value of $VPIF$ using the perception degrees. It transforms the perception degrees to perception values using a set of rules. The rules set the perception value to *db* for a perception degree below threshold S_{min}; value *uc* for perception degree above threshold S_{min} and below threshold S_{max}; *bl* for perception degree above threshold S_{max}. Note that S_{max} indicates how easy a user believes any information he/she receives–low S_{max} indicates the user easily believes and high S_{max} indicates the user hardly believes any information he/she receives. S_{min} indicates how easy a user disbelieves any information he/she receives.

$$S_k^d(t+1) = \begin{cases} db \text{ if } S_k^+(t+1) \leq S_{min} \\ uc \text{ if } S_{min} < S_k^+(t+1) < S_{max} \\ bl \text{ if } S_{max} \leq S_k^+(t+1) \end{cases} \tag{4}$$

Change of information. Let $VPIF$ change at time $t+1$. Equation 5 formulates the perception of node k regarding $VPIF_c$, which changes to *bl*, and the perception value of all the other network users about $VPIF_c$, which changes to *uk*. Note that when $VPIF$ changes, $VPIF$ regains its privacy because the previous piece of information known to the network users becomes obsolete.

$$S_k^c(t+1) = \begin{cases} uk\, k! = a \\ bl\, k = a \end{cases} \tag{5}$$

Private information could have one of two states: is changed, or did not change since the previous time step. Equation 6 formulates–using characteristic function–the perception value of node k considering the possibility of change of the disclosed information. If the information is changed, then the perception value of $VPIF$ is the one computed using Equation 5. Otherwise, it is the one computed using Equation 4. If we assume that information changes follow a binomial distribution, we compute the time of occurrence of the change using the formula $t \equiv (0\, mod\, (1/Pr^c))$. We use this assumption in Equation 6. (Equation 6 could be changed to consider any other function that models the occurrence of changes of information.)

$$S_k(t) = \begin{cases} S_k^c(t) \text{ if } t \equiv (0\, mod\, (1/Pr^c)) \\ S_k^d(t) \text{ if } t \not\equiv (0\, mod\, (1/Pr^c)) \end{cases} \tag{6}$$

The privacy degree of $VPIF$ at time t, based on the definition of Subsection 2.1, is the probability that the perception value of $VPIF$ by a user of the network selected at random is not bl. Equation 7 computes this probability as the proportion of users of the network who do not believe the true value of $VPIF$– i.e., either did not receive $VPIF$ or did not have the true value of $VPIF$, or are undecided about $VPIF$ at time step t.

$$Pr(t) = 1 - (\frac{(\sum_{j \in U}(S_j^d(t) = bl)) - 1}{N - 1}) \tag{7}$$

Dynamics of the structure of OSNs: The equations above are independent of the structure of the OSN and they apply to a static network–do not consider a dynamic network. This abstraction helps to keep the model simple. The simulation considers the structure of the OSN and the dynamics of the network.

4 Simulation of the dissemination of private information in an OSN

4.1 Simulation model

We simulate the evolution of the privacy degree of a piece of private information disclosed by a user in an OSN of 100 users. The OSN is simulated as a graph generated using Newman and Watts′s (Newman and Watts, 1999) algorithm for small-world networks. The algorithm is implemented as a method in the *Networkx* Python package (Hagberg, Schult, and Swart, 2012). Figure 1 shows a sample network generated by the algorithm.

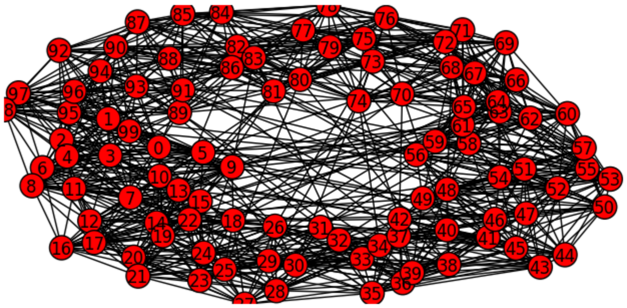

Figure 1 Example of the small-world network graph generated using Newman and Watts′s (Newman and Watts, 1999) algorithm

Table 2 summarizes the simulation parameters and Table 3 describes the distributions of the random variables that we use in the experiments.

At the beginning of the simulation, relationship strength and influence of neighbors are assigned to each edge in the graph, and an adoption of new

Table 2 Input parameters.

Variable	Description	Values
N	The number of nodes in the network.	100
NC	The neighborhood capacity for a node in the network.	20
Pr_{ae}	Probability for adding new edges in the network generation algorithm.	0.1
N_{it}	The number of time steps of the simulation.	30
S_{min}	The upper bound of perception degree required for a node to have its perception not be *db*.	0, -0.3, -0.7, -1
S_{max}	The lower bound of perception degree required for a node to have its perception be *bl*.	0.1, 0.4, 0.7, 1
D	Relationship strength threshold of $VPIF$ by any node to its neighbors.	0.1, 0.3, 0.5, 0.7, 0.9, 1
Pr_c	Change probability of $VPIF$.	0.1, 0.3, 0.5, 0.7, 0.9, 1

Table 3 Random variables and their statistical properties.

Random variable	Description	Value range	Statistical value distribution
Pr_i^{jk}	Influence of node j on node k.	0 to 1	Exponential distribution with mean 1/10
Pr_d^{jk}	Relationship strength of Node j on Node k.	0 to 1	Exponential distribution with mean 1/7
W_k	Weight of the sum of perception degrees of the neighbors of k in the new perception of k about $VPIF$.	0 to 1	Uniform distribution
$Init\_nds$	The selection of initial node which disseminates its own private information to its neighbors.	1..N	Uniform distribution

information weight is assigned to each node in the graph according to the distributions shown in table 3. Then, a user of the network, a *subject*, is chosen uniformly at random to disclose his/her private information, $VPIF$, to a subset of his/her immediate neighbors. At this point, the subject's perception value of $VPIF$ is set to *bl* and the perception value of all remaining users of the network is set to *uk*. At each subsequent iteration of the simulation (which represents a new time step) nodes that received $VPIF$ disseminate their perception values about $VPIF$ to their immediate neighbors. Relationship strength of each link affects the dissemination of $VPIF$.

The simulation keeps a global list of nodes who have received $VPIF$ as well as the perception value of each node of the network regarding $VPIF$. The list is updated at each iteration. Upon receiving perception values about $VPIF$ from his/her neighbors, each node computes his/her perception degree about *VPIF* using Equation 2 and Equation 3. Then, each node uses Equation 4 to decide whether to believe, not believe, or be undecided about $VPIF$.

While the information is being disseminated in the OSN, the subject could initiate the dissemination of a change of the information. Then the subject's perception value is set to bl and the perception values of all the other users are set to uk.

The experiments measure the evolution of the privacy degree of private information disseminated by a subject in the OSN. Equation 7 computes the privacy degree at iteration t.

4.2 Simulation results

This section reports about and analyzes the simulation results of the effects of relationship strength threshold, lower and upper bounds of perception degrees, change of information, and dynamics of the network on the privacy degree of disseminated information. The experiments were replicated 100 times and results are averaged.

4.2.1 Effect of relationship strength threshold on the evolution of privacy degree in a static OSN

In this experiment, we assign static values to S_{min}, S_{max}, and Pr^c and we vary D. (We consider that the relationship strength between the nodes is the same: D.) Figure 2 shows the effect of relationship strength threshold, D, on the privacy degree of $VPIF$. The figure shows that when D is between 0.5 and 1, the privacy degree is close to 1. However, the privacy degree declines and reaches a saturation point when D is between 0.3 and 0.1. For instance, the privacy degree gets close to 0 and continues at that level (but does not reach 0) for D equal to 0.1, at time step 3. The privacy degree declines and reaches a saturation point 0.6 at about time step 10 when D is 0.3.

We conclude that changing the relationship strength threshold strongly affects the privacy degree of $VPIF$. We also conclude that the privacy degree shows a decline over time until it reaches a saturation point; the saturation point depends on the relationship strength threshold.

4.2.2 Effect of lower and upper bounds of perception degree on the evolution of privacy degree in a static OSN

In this experiment, we assign static values to D, and Pr^c; and we vary the lower and upper bounds of perception degree, S_{Min} and S_{Max}. Recall that S_{Max} indicates how easy/difficult users believe new information they receive, which they adopt and disseminate further.

Figure 3 shows that for small values of S_{Max} the privacy degree of $VPIF$ drops dramatically to a saturation point in the first 3 time steps, but does not reach 0. While we also observe this behavior for S_{Max} values of 0.1 and 0.4, we see that S_{Max} value of 1 has almost perfect preservation of privacy because users disbelieve the information; therefore, they do not disseminate it.

Figure 4 shows the effect of S_{Min} on the evolution of privacy degree. As opposed to S_{Max}, which seems to have a huge impact on the privacy degree of

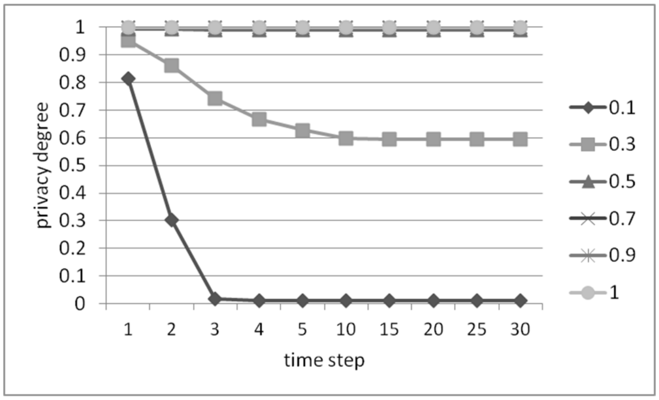

Figure 2 Effect of relationship strength threshold on privacy degree. Parameters: S_{Max}=0.1, S_{Min}=-0.7, Pr^c=0.

$VPIF$, S_{Min} only moderately affects the privacy degree, with lower S_{Min} values, leading to lower privacy degree.

We conclude that changing the upper bound of the perception degree, S_{Max}, strongly affects the privacy degree of $VPIF$. However, changing the lower bound of the perception degree, S_{Min}, has a small effect on the privacy degree of $VPIF$.

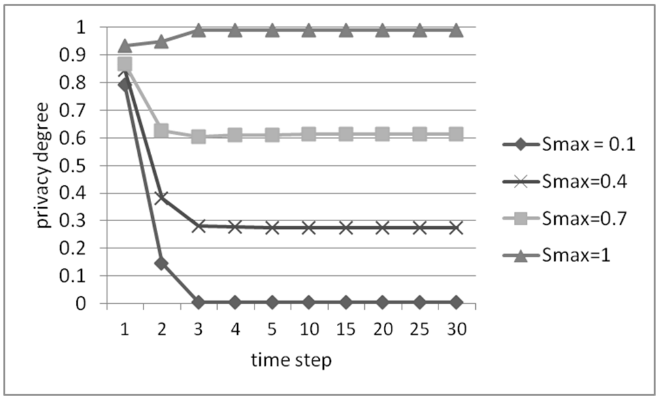

Figure 3 Effect of S_{Max} on privacy degree. Parameters: S_{Min}=-0.3, Pr^c=0 and D=0.1.

4.2.3 Effect of information change probability on the evolution of the privacy degree in a static OSN

In this experiment, we assign static values for S_{min}, S_{max}, and D, and we vary Pr^c of the disseminated $VPIF$. Figure 5 shows the effect of information change probability on the privacy degree. The figure shows that the privacy degree first declines to a level dependent on the change probability and gets close to 0 (It does not get close to 0 in the case of change probability 0.1 and reaches close to 0 for the other change probabilities). Then, when an information change occurs, the privacy

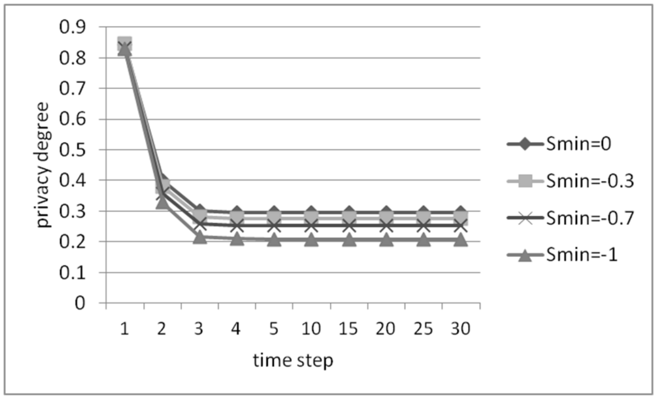

Figure 4 Effect of S_{Min} on privacy degree. S_{Max}=0.4, Pr^c=0 and D=0.1.

degree raises to a level close to 0.6. Finally, it declines again to reach a saturation point close to 0.2.

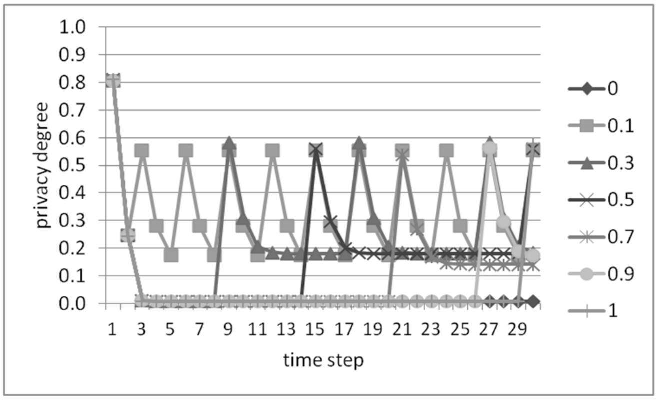

Figure 5 Effect of data change probability on privacy degree. Parameters: S_{Max}=0.1, S_{Min}=-0.3, and D=0.1.

4.2.4 Evolution of privacy in dynamic social networks

The addition of nodes and links (or their deletion) to an OSN affects the privacy degree of disseminated private information. In this experiment, we expand the OSN by one node at each time step–which corresponds to a growth of around 1% per time step, using a preferential attachment model for network growth (Barabasi and Albert, 1999).

Figure 6 shows a comparison of the evolution of privacy degree between static and dynamic networks for different information dissemination thresholds. The figure shows that addition of nodes to the network, for low dissemination thresholds, increases the privacy degree of disseminated information. We conclude that the growth of the network does not cause privacy loss for the cases considered here. Instead, it slightly increases the privacy degree of $VPIF$.

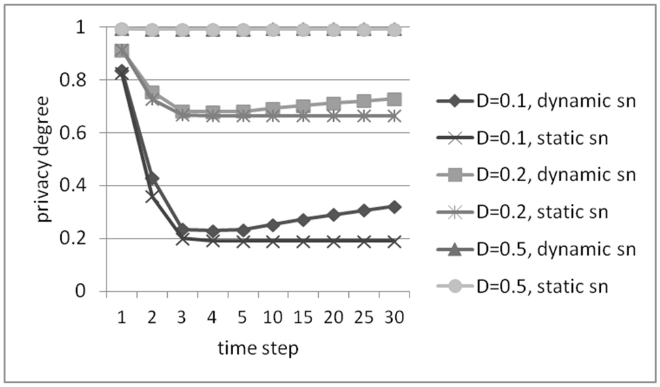

Figure 6 Evolution of privacy degree in dynamic vs. static social network. Parameters: S_{Min}=-0.3, S_{Max}=0.4, and Pr^c=0.

5 Related Work

Sensitive attribute inference problems in OSNs have been studied by researchers including Zheleva and Getoor (Zheleva and Getoor, 2009), who addressed the problem of inferences of private user attributes from public profile attributes, links, and group memberships in OSNs, whereas He et al. (He, Chu, and Liu, 2006) investigated the effect of social relations on sensitive attribute inference, and Becker and Chen (Becker and Chen, 2009) introduced PrivAware, a tool to measure private data inference risks in Facebook, where the privacy risk is attributed only to direct friend relationships.

More recent studies in social network privacy have focused on the loss of privacy of the profile data of users due to their social contacts and the network structure. Cutillo et al. (Cutillo, Molva, and Onen, 2011) analyzed the relationship between the social network graph topology and the achievable privacy. They observed that metrics such as the degree and the clustering coefficient of nodes severely affect user privacy with respect to identity/friendship privacy and usage control, while the mixing time of random walks in the social network graph plays an essential role in preserving the users′ communication untraceability. Anwar et al. (Anwar et al., 2009) proposed a privacy-preserving tool to enable a user to visualize the view that other users have of his or her Facebook profile, on the basis of the specified privacy policies. One of the first studies on measuring the amount of private information leaked from profile information on social networks was performed by Maximilien et al. (Maximilien et al., 2009). In their model, the privacy index of a user quantifies the user′s privacy risk caused by his privacy settings. The proposed model is based on a simple combination of two parameters: the sensitivity of a profile item, which depends on the nature of the item itself; and the visibility of the item, which captures how widely known the value of the item becomes in the social network. Gundecha et al. (Gundecha, Barbier, and Liu, 2011) demonstrated how much security an individual user can improve by unfriending a vulnerable friend. They also showed how security and privacy weakens if newly accepted friends are unguarded or unprotected. In their model, an individual is considered vulnerable if any friend in the network of friends has insufficient security and privacy settings to protect the entire network of friends. While these studies provide different ways of quantifying the privacy risk for individuals, their models are solely based on profile data, i.e. they fail to capture the privacy risks of other information about a user such as those extracted from wall/blog posts or information that changes with time.

One study that focuses on the privacy loss of non-profile data is that of Ngoc et al. (Ngoc et al., 2010; Kamiyama et al., 2010), which presents a metric to measure information leaked from blogs on social network sites, based on probability and entropy. Their proposed metric measures the density variation of probability distributions before and after adversaries attain helpful information from blogs; i.e., it measures the variation in the privacy value before and after information is published. In their model, the total privacy leaked from blog sentences means the change in the privacy value that is had by subtracting the privacy after sentences are posted from the privacy before the sentences are posted. In that work, the authors model the privacy loss of an event by considering joint information gathered from multiple blogs instead of considering the flow of information along the links in the social networks.

Most relevant to our work is the work of Carminati et al. (Carminati et al., 2011), who propose a probabilistic approach to estimate illegal leakage of resources in an OSN, where access control is regulated according to the topology-based paradigm. Specifically, they show how to compute the probability that a resource propagates from one user to another on the set of paths that link the two users. They quantify the Unauthorized Access Risk as an upper bound to the probability that sensitive resources reach any unauthorized user in an OSN that enforces topology-based access control. Another similar work is that of Wang et al. (Wang et al., 2011), who present a network-centric access control paradigm that explicitly accounts for the network effects in information flows. Their work is an attempt to study the impact of network effects (in socio-information networks) in risk-based access control. They use a multi-layer network model to capture subject-subject, object-object, and subject-object relationships, and encode relationships as intra-network or inter-network links. They apply a generic information flow model to quantify the qualification of a subject to access an object, where information is viewed as fluid that flows along links in socio-information networks. While their approach handles network evolution, it does not address the change in the privacy risk due to change of a data item.

There are also some experiments that aim to model information dissemination in social networks such as Liben-Nowell and Kleinberg's work (Liben-Nowell and Kleinberg, 2008). However, we are not aware of privacy models for information dissemination in an OSN that uses real data–many researchers are still working on the subject.

6 Conclusion and future work

OSNs exhibit the small-world property (short chain of nodes connects even the most distant users of the network) which leads to the belief that the privacy of information vanishes when the information is disclosed and disseminated in an OSN. Dissemination of information in an OSN exhibits a set of factors:

relationship strength, influence of neighbors, adoption of new information, change of information, and dynamics of the structure of OSNs. We believe that these factors affect the privacy degree of disseminated information in OSNs.

A user may disclose a set of private information to his/her friends, such as name, date of birth, mother's maiden name, etc. An identity thief could collect a set of appropriate information and use them to, for example, get a credit card on behalf of the person identified using the information (identification of a user using a set of private information is used in the USA and Canada, but not in Europe for example). In this work we investigated the effects of disclosure and dissemination of a single piece of private information in an OSN on the privacy degree of the information. In this paper, we propose a time series model for the evolution of the privacy of a piece of information disclosed in an OSN by its subject. We show, through simulating the dissemination of a private piece of information in an OSN considering the factors listed above, that the privacy of information does not vanish; it declines to a saturation point related to the factors of the information dissemination.

The conclusion that the privacy degree of information disseminated in an OSN does not vanish is not meant to encourage the disclosure of private information in an OSN. The model is especially expected to prove useful for models of prediction of privacy degree of disseminated information in an OSN conditions–represented by the factors.

The work shows how likely a user of an OSN gets the disseminated private information under varying values of several factors of information dissemination in OSNs. It uses thresholds to model the functions. Subsequent research questions include the following. Are there other factors that affect the privacy of disseminated private information in OSN? What are they, if any? What are the threshold values, and the weights used in the function? How good do the functions that we used–e.g., characteristic function for Equation 6, membership function for Equation 4, and weighed mean function for Equation 3–model the privacy degree of information disseminated in OSNs?

Future work will involve conducting experiments using real-world data crawled from publicly available portions of OSNs such as Facebook (Facebook, 2012) to see the fitness of the model for datasets with different characteristics. We will also investigate the effect of other factors such as deception–wrong information that a user may disclose in the network to deceive potential receivers.

References

Acquisti, Alessandro (2004). "Privacy in electronic commerce and the economics of immediate gratification". In: *Proceedings of the 5th ACM conference on Electronic commerce*. EC '04. New York, NY, USA: ACM, pp. 21–29. ISBN: 1-58113-771-0. DOI: 10.1145/988772.988777. URL: http://doi.acm.org/10.1145/988772.988777.

Anwar, M.M. et al. (Sept. 2009). "Visualizing Privacy Implications of Access Control Policies in Social Network Systems". In: *Proc. 4th International Workshop on Data Privacy Management (DPM'09)*. St Malo, France, pp. 106–120.

Barabasi, A.L. and R. Albert (1999). “Emergence of scaling in random networks”. In: *Science* 286.5439, pp. 509–512.

Becker, J. and H. Chen (May 2009). “Measuring Privacy Risk in Online Social Networks”. In: *Proc. Web 2.0 Security and Privacy (W2SP'09).* Oakland, CA.

Cambridge Advanced Learner's Dictionary. (2012). URL: `http://dictionary.cambridge.org/dictionary/british/perception_1?q=perception`.

Carminati, B. et al. (Feb. 2011). “A probability-based approach to modeling the risk of unauthorized propagation of information in on-line social networks”. In: *Proc. 1st ACM Conference on Data and Application Security and Privacy (CODASPY'11).* San Antonio, TX, pp. 51–62.

Cutillo, L.A., R. Molva, and M. Onen (Dec. 2011). “Analysis of Privacy in Online Social Networks from the Graph Theory Perspective”. In: *Proc. Global Telecommunications Conference (GLOBECOM 2011).* Houston, TX, pp. 1–5.

Facebook (2012). *Facebook.* URL: `http://www.facebook.com/`.

Govani, Tabreez and Harriet Pashley (2012). *Student awareness of the privacy implications when using Facebook. Carnegie Mellon.* URL: `http://lorrie.cranor.org/courses/fa05/tubzhlp.pdf`.

Gross, Ralph and Alessandro Acquisti (2005). “Information revelation and privacy in online social networks”. In: *Proceedings of the 2005 ACM workshop on Privacy in the electronic society.* WPES '05. Alexandria, VA, USA: ACM, pp. 71–80. ISBN: 1-59593-228-3. DOI: `10.1145/1102199.1102214`. URL: `http://doi.acm.org/10.1145/1102199.1102214`.

Gundecha, P., G. Barbier, and H. Liu (Aug. 2011). “Exploiting vulnerability to secure user privacy on a social networking site”. In: *Proc. ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD'11).* San Diego, CA, pp. 511–519.

Hagberg, Aric, Dan Schult, and Pieter Swart (2012). *Networkx 1.6.* URL: `http://networkx.lanl.gov/`.

He, J., W. W. Chu, and Z. Liu (2006). “Inferring privacy information from social networks”. In: *Proc. IEEE International Conference on Intelligence and Security Informatics.* San Diego, CA, pp. 154–165.

Kamiyama, K. et al. (Oct. 2010). “Unified Metric for Measuring Anonymity and Privacy with Application to Online Social Network”. In: *Proc. International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2010).* Darmstadt, Germany, pp. 506–509.

Kempe, D., J. Kleinberg, and E. Tardos (Aug. 2003). “Maximizing the spread of influence through a social network”. In: *Proc. ACM International Conference on Knowledge Discovery and Data Mining (KDD'03).* Washington, DC, pp. 137–146.

Liben-Nowell, David and Jon Kleinberg (2008). “Tracing information flow on a global scale using Internet chain-letter data”. In: *Proceedings of The National Academy of Sciences (PNAS)* 105.12, pp. 4633–4638. DOI: `10.1073/pnas.0708471105`.

Maximilien, M.E. et al. (May 2009). “Privacy-as-a-Service: Models, Algorithms, and Results on the Facebook Platform”. In: *Proc. Web 2.0 Security and Privacy (W2SP'09).* Oakland, CA.

Mertsalov, K., M. Magdon-Ismail, and M. Goldberg (July 2009). “Models of Communication Dynamics for Simulation of Information Diffusion”. In: *Proc.*

International Conference on Advances in Social Network Analysis and Mining (ASONAM'09). Athens, Greece, pp. 194–199.

Milgram, Stanley (1967). "The small word problem". In: *Psychology Today* 1.1, pp. 61–67.

Mislove, Alan et al. (2007). "Measurement and analysis of online social networks". In: *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement.* IMC '07. San Diego, California, USA: ACM, pp. 29–42. ISBN: 978-1-59593-908-1. DOI: `10.1145/1298306.1298311`. URL: `http://doi.acm.org/10.1145/1298306.1298311`.

Newman, M. E. J. (July 2002). "The spread of epidemic disease on networks". In: *Physics Review E* 66, p. 016128.

Newman, M. E. J. and D. J. Watts (1999). "Renormalization Group Analysis of the Small-World Network Model". In: *Physics Letters A* 263, pp. 341–346.

Ngoc, T.H. et al. (Apr. 2010). "New Approach to Quantification of Privacy on Social Network Sites". In: *Proc. International Conference on Advanced Information Networking and Applications (AINA'10).* Perth, Australia, pp. 556–564.

Pinterest (2012). *Pinterest.* URL: `http://pinterest.com/`.

Rosen, Jeffery (2010). *The Web Means the End of Forgetting.* The New York Times. Published: July 21, 2010. URL: `http://www.nytimes.com/2010/07/25/magazine/25privacy-t2.html?_r=3&pagewanted=1&ref=technology`.

Shirey, R. (Aug. 2007). *Internet Security Glossary, Version 2.* RFC 4949 (Informational). URL: `http://www.ietf.org/rfc/rfc4949.txt`.

Wang, T. et al. (June 2011). "Modeling data flow in socio-information networks: a risk estimation approach". In: *Proc. ACM Symposium on Access Control Models and Technologies (SACMAT'11).* Innsbruck, Austria, pp. 113–122.

Watts, D.J. and S.H. Strogatz (June 1998). "Collective dynamics of small-world networks". In: *Nature* 393.6684, pp. 440–442.

Westin, Alain (1967). *Privacy and freedom.* New York: Atheneum.

YouTube (2012). *YouTube.* URL: `http://www.youtube.com/`.

Zheleva, E. and L. Getoor (Apr. 2009). "To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles". In: *Proc. World Wide Web Conference (WWW'09).* Madrid, Spain, pp. 531–540.