# Privacy scoring of social network users as a service

Vidyalakshmi B. S.
University of New South Wales
and CSIRO, Sydney, Australia

Raymond K. Wong
University of New South Wales
and National ICT Australia

Chi-Hung Chi
CSIRO
Tasmania, Australia

***Abstract*—Privacy oriented communication in social network has been researched extensively due to the dangers of private and personally identifiable information falling into wrong hands. To assist users better manage information disclosed, social network providers have introduced more controls like groups, lists and circles. With multitude of privacy controls on each social network, users are burdened with setting and managing privacy, on each social network separately as the friends grouping cannot be shared between social networks. This necessitates a service model which can be used across social networks, making use of information from all social networks. With the current privacy settings on social networks, users attitude towards privacy and communication cannot be accounted for although it has major influence on who sees what information. We propose to develop a framework for calculating privacy scores of friends from individual ego users' perspective. We contend privacy score would assist user in assessing his or her information sharing behavior and take an informed decision on who sees what information. Privacy scoring considers users' personal attitude i.e., disposition towards privacy and communication information. Privacy scores are estimated based on privacy scoring function using cubic bezier curve. Our experiments highlight the effective working of the proposed framework in estimating friends privacy scores with various ego users' dispositions to privacy and communication.**

## I. Introduction

Participation on social networks has surpassed all other online activities to be one of the top internet activity[1]. With increased online presence and information sharing comes the risk of information reaching more than the intended audience. This has warranted a need to protect our privacy online, just like the real world. To cater to the needs of online privacy, *Privacy as a service* [11], [18], [27] is drawing researchers and industrialists attention alike.

*Privacy* in social networks is an oxymoron as the very reason for being on social network is to articulate and communicate with extended social network of *friends*. The problem is not so much in disclosing information on social networks. In fact, early studies into social networks privacy [10], [26] have shown users discern that certain amount of information revelation as necessary to make social networks useful, *'Why have a profile if your profile does not say enough about who you are?'* [26]. The problem arises in disclosing the information to all instead of being directed at certain people.

Studies on privacy awareness highlight users' being aware of privacy and its implications [15], [26]. They try to navigate around privacy through many tactics like wall cleaning, aliases, proxies, friending behaviors among many [9] [21] [5].

To cater to this increased awareness and to make privacy as a priority, social network providers have introduced privacy controls to a granular level, being able to control visibility of each profile item and post. To support user further, *group* based and *list*[2] based communication was introduced along with continued support for privacy control through Friends and Friends of Friends (FOAF) methodology.

Granular controls vary from one social network to the next, are time consuming to set visibility forcing the user to choose between no visibility or going with the default recommended settings [1][3][4]. On the other hand, automatically created groups or lists are shown to be rigid, noisy and problematic in managing security [13] with users not utilising the feature[5], thereby bringing to the fore, a need for on-demand group creation.

Community extraction and clustering have been proposed as ways to create groups for a user. Community extraction using network graph has been extensively studied by [8], [20], [6] and has been particularly tested in social networks community detection in [13]. Jones et. al [13] try to answer the common criterias considered by users when creating groups through a survey of Facebook users while Eslami et. al evaluate disjoint clustering, overlapping clustering and hierarchical clustering algorithms for grouping friends. Utilising shared features of friends to learn, an active machine learning based grouping, *Regroup*, was proposed by [3]. Regroup, brings the suggested friends to the top, thereby assisting user in grouping friends on-demand.

Even though community extraction, clustering and context based grouping discussed above provides a powerful and best among the existing methods for automatic group creation, they suffer multiple problems some of which are discussed in [6] -

- Shaky ground truth for comparison. Groups extracted from existing user groups are not modified often to update changes, with one study showing nearly 80% of users having updated their friend lists only once [16]
- Complexity in defining contexts
- Friends often share multiple roles (or contexts) with the ego user and are a part of composite hierarchical structure in social settings. A university friend may share university

[1]http://www.businessinsider.com.au/social-media-engagement-statistics-2013-12

[2]https://www.facebook.com/notes/facebook/improved-friend-lists/10150278932602131

[3]http://www.adweek.com/socialtimes/facebooks-complex-privacy-settings/307603

[4]http://techcrunch.com/2014/04/08/facebook-privacy-settings/

[5]http://techcrunch.com/2010/08/26/facebook-friend-lists/

978-1-4673-7281-7/15 $31.00 © 2015 IEEE
DOI 10.1109/SCC.2015.38

context along with specific course context with the ego user
- Uncertainty in grouping. Users often face indecisiveness in deciding the boundary conditions for a group. With friends sharing multiple roles, the problem is exasperated

As such, there is a need for a privacy quantifying mechanism that can be used to 1) group users 2) take informed decision in distributing information in social networks. The mechanism should be intuitive to the end user while taking into perspective users individual attitudes towards privacy in social network. In this paper we propose to shift focus from existing traditional model of communication and build a personalised and perceptive way of communication by utilising ego users' disposition to privacy and communication. We propose a framework for estimating ego users' friends privacy orientation or privacy sensitivity as a *Privacy Score* from ego users' perspective. With user having to comprehend ever increasing privacy settings in managing his information and of doing so on multiple social networks with diverse and non-overlapping privacy controls, it would be worthwhile in providing calculation of privacy score as a service where in one score can be used across all social networks.

Disposition to privacy is the inherent nature of an individual user. Users disposition influences how often he communicates, amount of information disclosed and communication frequency, communication with close friends alone or with acquaintances too, just like an extrovert communicating more than an introvert. We do not restrict users disposition to be classified as introvert or extrovert but provide a scale between 0 and 1 in defining their disposition towards privacy and communication.

Privacy score framework proposed uses cubic bezier curve in deriving a privacy scoring function. A cubic bezier in its parametric form, utilises the friends privacy sensitivity towards ego users information, defined by the FACT function to arrive at a privacy score. The purpose of our experiments is to show the working of the proposed model and the distribution of privacy scores of friends against varying dispositions to privacy and communication of the ego user.

With the Privacy scoring as a service, we can achieve the following:

- Privacy scoring of friends that is intuitive and able to adjust to non-varying dynamics of ego users' attitude towards privacy and communication
- A service that can efficiently utilise information from all social networks when ego user and friend are connected through multiple social networks
- Category based information dispersal - Based on sensitivity, information can be categorised with friends grouped based on privacy score with a direct relationship between sensitive category access granted to friends with acceptable privacy scores. Category based communication can be an alternative to the current group and list based communication.
- Privacy setting recommendation - Lists and groups currently available in social networks could be analysed for having friends whose privacy scores show *group invariance*
- Social study - The privacy scores and ego users disposition to privacy could be used to study the behaviors of social network users
- Common score across multiple social networks giving a single source of truth instead of lists, groups that need to be created across multiple social networks and cannot be shared among each other

Rest of the paper is organised as follows. We discuss the related work in Section II, define the problem being addressed in this paper in Section III, discuss why the proposed framework is better suited as a service in Section IV and submit the preliminaries for the proposed model in Section V. Proposed model is described in Section VI and experiments in Section VII. Section VIII concludes and highlights the directions for extension of the current work in the future.

## II. Related work

Privacy in social networks is an extensively researched subject. Classifying people according to their privacy concern has been studied much before the introduction of social networks [30], [31] with people divided into three broad categories: Privacy fundamentalists, Pragmatists and Unconcerned.

We briefly discuss previous works on privacy scoring in social networks.

A framework based on Item response theory (IRT) to estimate privacy scores of ego user has been proposed by Liu et. al [17]. Privacy score is assigned to each user based on sensitivity and visibility of information shared with friends. Privacy score of a user calculated using only profile items leaves out utilising posts, likes, comments, wall posts, thereby giving an incomplete picture of the actual privacy behavior of the user in social network.

Privacy quotient and Privacy Armor models [24] are proposed as a way to measure privacy leaks using the IRT model. The authors identify the importance of posts, tweets and other unstructured information that are the sources of privacy leakage. Privacy Armor model proposed, measures average privacy quotient of the group members, who would be probable audience of a message about to be posted. If the score does not meet the desired score, user is alerted who can then take appropriate further steps. While using posts, tweets and other unstructured data, information is classified as sensitive and not-sensitive based on the presence of profile items alone thereby restricting the leakages to 11 profile items discussed in the paper.

Another work utilising the IRT model [23] and extending the work of [17], propose to utilise background knowledge about the ego user from multiple social networks and auxiliary background knowledge available through blogs, comments. Background knowledge by themselves may not pose privacy risk but when combined with other information, it might disclose sensitive information. By collecting background knowledge and processing them through natural language processing

techniques, authors contend the sensitivity of the items to be closer to real sensitivity. Though the work identifies important aspects that need to be considered while assigning sensitivity scores to items, the method may be applicable only for certain attributes like religion and political orientation which are discussed in open forums from where auxiliary information can be scraped.

Privacy Index (PIDX) proposed [19], is a measure of a user's privacy exposure in a social network. PIDX is a numerical value between 0 and 100 with high value indicating high privacy risk in social networks. PIDX is the summation of privacy impact factors of each attribute visible where in each attributes privacy impact factor is the ratio of its privacy impact to full privacy disclosure. Extending their model, the authors have proposed [29], which measures the privacy exposure between any two given users i and j,$PIDX(i, j)$. The paper does not clearly mention the methods used for deep web searching and data aggregation. Also, the backbone of the proposed Privacy Index, the privacy impact factor is a static measure focusing on few attributes listed in the paper.

All the above papers score ego user using the information he has disclosed to his friends. Different than these papers, in this paper we propose to estimate privacy scores of friends from an ego users' perspective utilising ego users disposition to privacy and communication. The friends are restricted to the ego-subnetwork of the ego user.

## III. Problem Abstraction

In a social network, an ego network is the subnetwork of a user called the *ego user* and friends connected to the ego user. The ego network consists of connections between the ego user and his friends and among his friends. Figure 1 shows an ego network with $U_1$, the ego user who is connected to his friends $f_1$,$f_2$,$f_3$,$f_4$ and $f_5$. In this paper, we specifically use *friends* to refer to all friends of the ego user in a given social network.

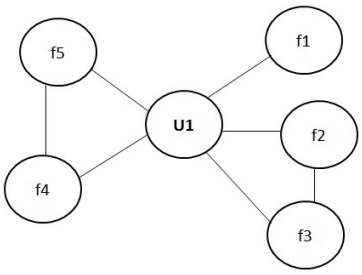


Fig. 1. Sample Ego Network

The proposed model aims to assign privacy scores to all friends from ego user's perspective, utilising ego users' disposition to privacy and communication. Each ego user utilises the service to assign a privacy score to each of his friends within a range of 0 (maximum privacy) to privacy threshold $T^0$ (low privacy), as set by the ego user. Higher the privacy score, lower the privacy. We model privacy scoring using cubic bezier curve, a mathematical function that accommodates the personality of the ego user.

## IV. Why privacy scoring service

According to the latest statistics, 52% adults use multiple social networking sites[6]. Users are connected to some friends on more than one social network while some friends are specifically connected to the ego user through only one social network. The commonality, ego user and friends share, across social networks is their disposition to privacy and communication which is an inherent nature to any individual. A service model will be able to utilise the commonality in disposition between social networks as also utilising communication with common friends across social networks. With privacy scoring as a service, we can -

- Across multiple social networks, common disposition to privacy and communication can be utilised in assigning privacy scores. The service can also cater to an ego user who would like to have different dispositions on each social network
- Communication data can be pooled from across multiple social networks in finding the percentage of communication directed at a friend in a given period of time, required for calculating C as described in Section V-A
- Implementation complexity of the privacy scoring model, bezier curve function, can be masked from the end users

Hence, we contend that privacy scoring implemented as a service has advantages over scoring incorporated individually into each social network.

### A. Service Implementation

This Section briefly describes the implementation of privacy as a service. Architecture diagram, shown in Figure 2, is intended in highlighting the major blocks that form privacy scoring as a service. It is notable that by keeping the entity *privacy scoring framework* separate, it is easier for updating the model with enhancements. Social network users using the privacy scoring service can maintain consistent disposition to privacy and communication across all social networks they use. Ego user provides input to the service from a particular social network, receiving in return the privacy scores of all friends in that social network, from privacy scoring service.

Figure 3, a control flow diagram, describes the interaction between ego user and privacy scoring service at a high level. Service receives as input $l^0$,$h^0$, $T^0$ and the social network for which privacy scoring is to be done. Sorted friend list will continue to be static once initialised except where there is an addition of a new friend. Consider a situation where ego user acquires more knowledge about a friend who is indulging in non-privacy enhancing practices. In such a situation, sorted friends list may need to be modified too. It is notable that users can be grouped into pre-defined groups rather than sorting each and every friend as described in Section V-B.

## V. Background

Attitude of a user in a social setting has been extensively researched. In social-psychology literature, values, beliefs and

[6]http://www.pewinternet.org/2015/01/09/social-media-update-2014/

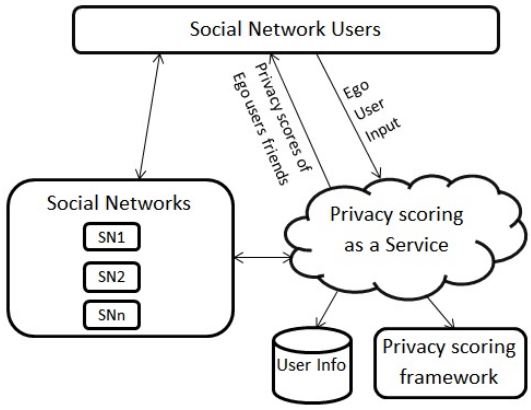


Fig. 2. Architecture Diagram

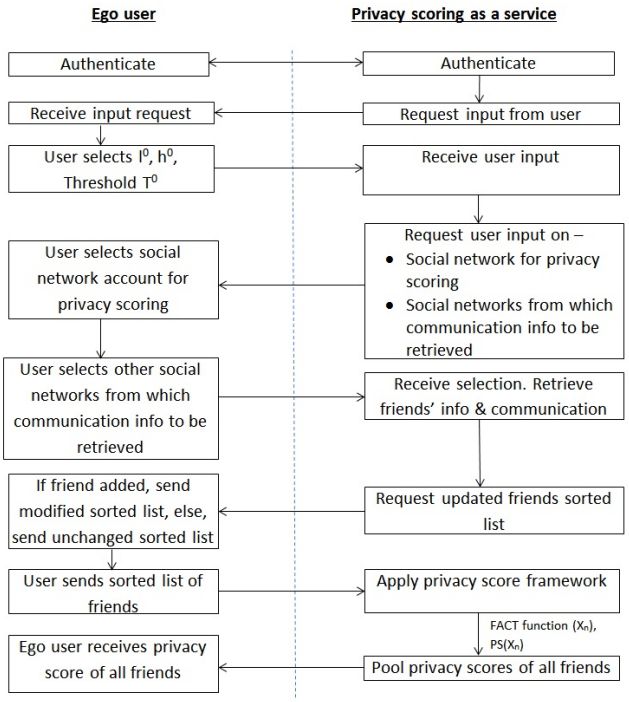


Fig. 3. Control Flow Diagram

attitudes are connected to each other, empirically and theoretically [2]. Since, Online social network is a social platform with problems and practices exhibited online and offline being entwined and shared [4], we can borrow from existing studies. In this work, we conceptualize 'privacy' as an attitude, defined as a learned predisposition to respond in a favourable or unfavourable manner to information disclosure [7].

Due to the complexity and inconsistencies of defining and measuring privacy per se, *privacy concerns* has been used as a proxy in its place, by many studies.

In this work, we use privacy scoring to measure privacy concerns of an *Ego user* towards his *friends* in social network. Attitude (Attitude and Disposition are used interchangeably) of the ego user towards privacy and communication is employed in arriving at a privacy score. A cubic bezier curve function calculates privacy values of friends, utilizing ego users' disposition. Ego users' predisposition towards his *friend's attitude towards privacy* or *How privacy aware is ego users' friend towards ego users' information* is also accounted for (as $P_n$) in estimating privacy score in the Friend_Attitude_Calculator function. We assume that the attitude of friends towards ego users' information does not change drastically.

### A. Disposition to privacy and communication

Disposition to privacy is the inherent attitude of an individual towards privacy of his or any other information under his possession. This attitude is neither dependent on any individual nor on circumstances but can change with time, albeit slowly.

In the proposed model, disposition to privacy is denoted by $l^0$, which is bounded between 0 and 1. Each ego user is free to choose his disposition towards privacy from $l^0 = 0$, representing lax privacy orientation through $l^0 = 1$, utmost privacy concerned behavior.

Disposition to communication is the ego users' attitude towards communication. An ego user may be communication oriented individual who is on social network to communicate and gain social capital [5], [25]. An ego user could also be a silent observer who is just consuming information from others rather than giving out any [28]. This attitude of an ego user is very different from his disposition towards privacy and hence needs to be accommodated as a separate variable that contributes to privacy score calculation. Disposition to communication is the inherent attitude of an ego user towards communication online, denoted as $h^0$, with value between 0 and 1. An Ego user with $h^0 = 0.1$ denotes a user who is very communication oriented while $h^0 = 0.9$ denotes a user who is mostly a silent observer and communicates with a select few.

Based on ego users' disposition to privacy $l^0$ and disposition to communication $h^0$, Privacy scores assigned to friends varies.

### B. Friend_Attitude_Calculator (*FACT*) function

Each users' perception of his social network friends is based on personal experiences with the other person. It is therefore, imperative that ego user is the best judge in deciding who is more privacy oriented or less privacy oriented among his friends. It is intuitive for the ego user to judge *friend f1 is more privacy oriented than f2*. We propose to use this notion of evaluating friends in relation to their perceived privacy orientation in the eyes of the ego user. Friends are sorted by the ego user according to their privacy orientation towards ego users' information. The sorted list of friends is used in determining the overall attitude of the ego user towards his friends. $P_n$ denotes each friends position in the sorted list. It should be noted that this sorting is one-off, and will change only if ego user discovers friends' privacy behavior contrary to his current understanding.

Average number of friends in popular social networks is more than 100 (Eg., Half of Facebook users have more than 200 friends[7]). We are aware that sorting these many friends would be cumbersome for an established (A user not new to social network) user. To overcome this, ego user could add friends into sorted buckets and consider sorted buckets as the list of sorted friends.

Communication in social network is bidirectional. The ego user can view his friends profile and posts while his friends can view ego users profile and posts. While the communication is bidirectional, we are interested in the information communicated from ego user to his friends. In the proposed model, profile items, posts, comments from ego user to his friends is considered in calculating the total communication.

Frequency of communication over a period of time with friends indicates the closeness of the ego user and friends. Interaction frequency has been studied as good indicator for tie-strength between users in social network [14], [32], [12]. We make use of this to calculate communication frequency with each friend's communication $C_n$ as a percentage of total communication. Communication is denoted by **C** and has a value between 0 and 1.

**F**riend_**A**ttitude_**C**alculator (**FACT**) function is calculated as in Equation 1:

$$X_n = (\frac{P_n}{t_x - 1}) * (1 - C_n) * (t_x - 1) \quad (1)$$

where $X_n$ denotes the output for friend $n$, $P_n$ is the position of user in the sorted list of friends, $C_n$ denotes communication percentage and $t_x$ = (Total Number of friends) +1. $X_n$ is further used in privacy score calculation of friend $n$ as shown in equation 3. The output of the FACT function is a representation of friends from ego users' perspective.

### C. Cubic Bezier Curve explained

Bezier curves are a form of parametric function to draw a smooth curve. The curve starts at a point $P_1$, going through points $P_2, P_3 ... P_{n-1}$ and terminating at $P_n$. Bezier curves are defined by the control points that control the shape of the curve. As shown in Figure 4, by fixing the positions of $P_1$ and $P_4$, and moving the points $P_2$ and $P_3$ inside the defined rectangle, we can adjust the curvature of the curve. Such a curve formed by four control point is called a cubic bezier curve.

A parametric bezier curve is a curve that is determined by coordinate pairs of (x,y) points, in which x and y values are determined by a separate variable, time 't' with values between 0 and 1. The equation for cubic bezier curve in terms of 't' is given below.

$$B(t) = (1-t)^3 P_1 + 3(1-t)^2 t P_2 + 3(1-t) t^2 P_3 + t^3 P_4 \ , \quad t \in [0,1] \quad (2)$$

[7]http://www.pewresearch.org/fact-tank/2014/02/03/6-new-facts-about-facebook/

We make use of cubic bezier curve in privacy score function due to its flexibility in plotting geometric curves. In adopting bezier curves to privacy scoring function, we determine the ordinate as a function of abscissa instead of both ordinate and abscissa determined by the temporal variable 't'. The output of FACT function, determines the abscissa. Privacy scoring function takes abscissa $x$, and returns the corresponding ordinate ($y$), the "Privacy Score" of a node (friend) in the ego network.

## VI. Model

### A. Privacy scoring function

Disposition to Trust has been used by [22] to arrive at Trust scores in a peer-to-peer network. We are influenced by this work in arriving at a privacy scoring framework. We differ from [22] as we use 1) Disposition to privacy and disposition to communication, two behaviors or attitudes of the user in a social network as opposed to disposition to trust alone and we use 2) Cubic bezier curve in accommodating the behaviors instead of Quadratic bezier curves.

Adopting cubic bezier curve, we define a mathematical function, **P**rivacy **S**coring function ($PS(X_n)$) to arrive at privacy score of each friend of the ego user.

The input $X_n$ is calculated through FACT function and is a positive real number between 0 and $t_x$ +1. $t_x$ is the number of friends of the ego user. The output of the $PS(X_n)$ represents the corresponding privacy score of friend $n$. Privacy threshold, denoting the maximum privacy score $T^0$ = $t_y$ can be independently fixed by the end user.

$PS(X_n)$ function implemented using cubic bezier curve, is defined by four control points. By fixing the first and last points P1 and P4, we end up having two control points P2 and P3, each dictating the ego users disposition towards privacy and communication, respectively. The four points of cubic bezier curve that are used in privacy scoring function are:

- The origin point $P_1(0,0)$
- The privacy personality point $P_2(l_x, l_y)$
- The communication personality point $P_3(h_x, h_y)$
- The threshold point $P_4(t_x, t_y)$

The Privacy scoring function $PS(X_n)$ passes through origin (P1) and threshold (P4) points. Privacy personality (P2) and communication personality(P3) points dictate the curvature of the curve. The curve is completely contained in the convex hull (Outer rectangle shown in Figures 4 and 5) of its control points.

We assume that it is sufficient to move the personality points P2 and P3 through the second diagonal represented by $l_x = \frac{-t_y}{t_x} * l_y + t_y$ and $h_x = \frac{-t_y}{t_x} * h_y + t_y$ of the formed rectangle, to plot a large panel of personalities of ego users. It is notable that, privacy scoring function is monotonic, i.e., since, the points P2 and P3 are restricted to trace the second diagonal of the defined rectangle alone, the ordinate yields a single value for any abscissa value.

P2 and P3 are obtained using disposition to privacy $l^0$ and communication $h^0$, respectively. P2, P3 and Privacy Score function are derived as follows:

$$PS_{l^0,h^0,t_x,t_y}(X_n) = (3l_y - 3h_y + t_y)(\propto X_n)^3 + 3(h_y - 2l_y)(\propto X_n)^2 + 3(\propto X_n)l_y \quad (3)$$

where,
$(\propto X_n) = (3l_x - 3h_x + t_x)(X_n)^3 + 3(h_x - 2l_x)(X_n)^2 + 3X_n l_x$,
$0 \leq l_x \leq t_x$ && $t_x > 0$,
$0 \leq h_x \leq t_x$ && $t_x > 0$,
$l_x = (1 - l^0) * t_x, l_y = t_y * l^0$
$h_x = (1 - h^0) * t_x, h_y = t_y * h^0$

### B. Privacy scores generation

The points P2$(l_x, l_y)$ and P3$(h_x, h_y)$ take values based on the ego users' disposition to privacy ($l^0$) and disposition to communication ($h^0$), respectively. Threshold $T^0$ = $t_y$ determines the maximum privacy score while $t_x$ = Number of friends +1.

As given in Figure 4, the curve will be smooth when $l^0$ and $h^0$ have same values or values that are very near to each other. As given in Figure 5, the curve will ease in and then ease out with inflection points when $l^0$ and $h^0$ values are inversely proportional to each other (eg.,$l^0$=0.7 and $h^0$=0.3). Figure 4 and 5 have been plotted with $t_x$ = 6 and $t_y$ = 50.

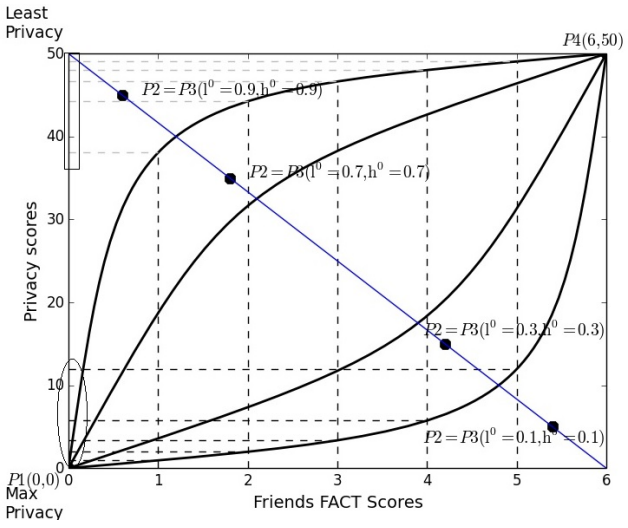


Fig. 4. Privacy Scoring where $l^0$ and $h^0$ are directly proportional

Figure 4 highlights the various personalities of ego users as classified by [30], [31] with people divided into three broad categories: Privacy fundamentalists, Pragmatists and Unconcerned.

Privacy scores of friends as in Figure 4 (oval on y-axis), is nearer to privacy score 0 (max privacy) when $l^0$ and $h^0$ are 0.1, indicating personality of an ego user who would most likely treat all friends as same and has his information, profiles and posts, shared publicly. He has very little regard for privacy and is *At risk* individual.

Privacy scores are towards the maximum threshold (rectangle on y-axis) for an ego user who displays high disposition

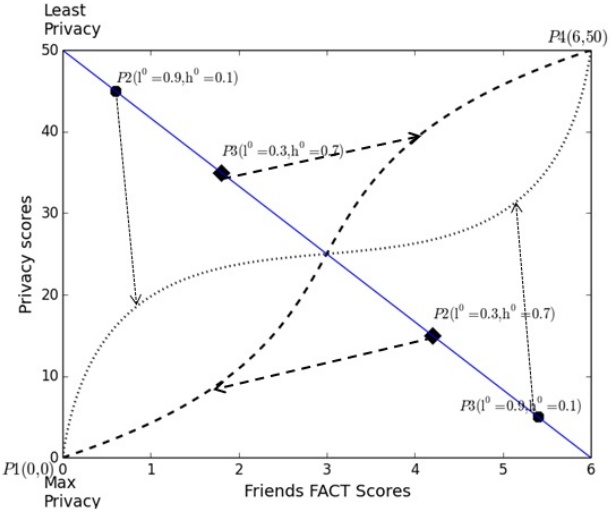


Fig. 5. Privacy Scoring where $l^0$ and $h^0$ are inversely proportional

towards privacy and communication (eg., $l^0 = h^0 = 0.9$). The ego user displaying this personality is very privacy oriented and is likely a silent observer, communicating only with trusted few. Privacy scores of friends are concentrated towards the threshold (low privacy).

Communication disposition and privacy disposition are directly related, in general. Our proposed framework works for cases where ego users select $l^0$ value to be inversely proportional to $h^0$ value. Figure 5 shows the behavior of Privacy scoring function in such situations. An ego user who knows his audience and wants to utilise social network to communicate with his friends at the same time keeping out those whom he considers less privacy oriented, exhibits these dispositions.

## VII. Experiments

We start by giving a brief description of the data, used in the experiments.

### A. Synthetic dataset creation

According to a latest study, half of Facebook users have more than 200 friends with the case being no different with other popular social networks. In the generated synthetic dataset, we consider an ego user with 200 friends, $f_1$ to $f_{200}$, with their positional number $P_n$ (1 to 200), sorted ascending. Each friends communication, $C_n$, is used in privacy scoring and is calculated as a percentage of the total communication. Communication is generated uniformly at random from intervals [0,1]. It is important to note that a single post can be received by multiple people with group communication at play, communication percentage of all friends put together will be greater than the total communication. In the general setting, communication is frequent with a close friend, whom ego user considers more privacy oriented (Low $P$ number). But, instead of having communication as in general setting, we have

generated communication at random and plotted the privacy scores so as to bring out the ability of the proposed framework to work even if the communication is erratic. We contend that, since the communication $C_n$ is used as a percentage of the total communication, synthetic dataset will yield similar results as the real data gathered from social networks.

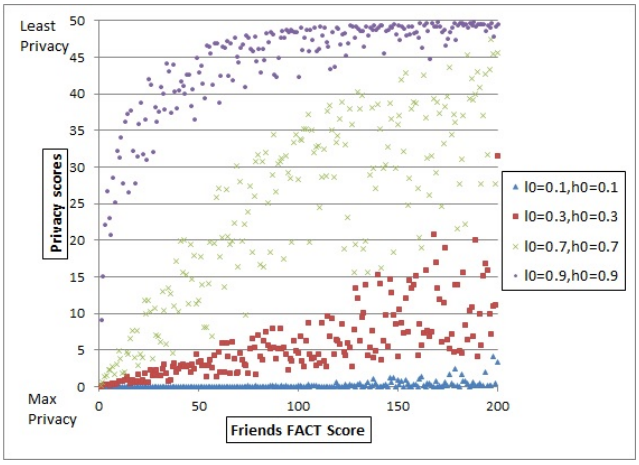


Fig. 6. Experiment results with varying $l^0$ and $h^0$

For the purpose of the experiments, we fixed the privacy score threshold, $T^0 = 50$. We plot different attitudes of an ego user with varying $l^0$ and $h^0$ values, keeping $P_n$ and $C_n$, constant. Figure 6 shows the privacy scores of all 200 friends with varying $l^0$ and $h^0$ values. It is important to note that lower the privacy score, higher the privacy orientation (or privacy sensitivity) of the friend $f_n$.

As observed from Figure 6, with higher privacy and communication disposition (($l^0 = 0.9$,$h^0 = 0.9$) and ($l^0 = 0.7$,$h^0 = 0.7$)) privacy scores of friends are concentrated in the higher stratas, towards the Threshold $T^0$ (low privacy), while with lower disposition (($l^0 = 0.3$,$h^0 = 0.3$) and ($l^0 = 0.1$,$h^0 = 0.1$)), scores are concentrated nearer to 0 (max privacy).

**Example:** Let's consider four ego users, with varying dispositions towards privacy and communication as listed in the header of table 1 and as plotted in Figure 6. By applying the $PS(X_n)$ function, privacy scores of the friends are calculated. Table 1 gives a snapshot of the Privacy scores of 200 friends from Figure 6 in positions $P_n$ with communication percentage $C_n$.

### *B. Naive Privacy score*

The purpose of this Section is to describe a simple way of calculating privacy scores and highlight the disadvantages of Naive approach against the proposed framework.

Communication activity has been shown to be a good indicator of closeness or relationship strength between users [32], [27], [12]. Intuitively, higher the communication frequency, $C_n$, closer (viz., close friend) the friend $n$ is to ego user and lower their $P_n$ positional number. Considering the threshold

| $P_n$ | $C_n$ | $l^0 = h^0$ =0.9 | $l^0 = h^0$ =0.7 | $l^0 = h^0$ =0.3 | $l^0 = h^0$ =0.1 |
|---|---|---|---|---|---|
| 7 | 0.79 | 3.22 | 0.84 | 0.15 | 0.04 |
| 39 | 0.45 | 32.33 | 12.34 | 2.30 | 0.60 |
| 66 | 0.60 | 35.15 | 15.00 | 2.82 | 0.74 |
| 81 | 0.67 | 35.46 | 15.35 | 2.89 | 0.75 |
| 111 | 0.79 | 33.52 | 13.37 | 2.50 | 0.65 |
| 134 | 0.28 | 46.41 | 37.60 | 11.11 | 3.15 |
| 159 | 0.45 | 45.93 | 36.21 | 9.97 | 2.78 |
| 172 | 0.51 | 45.69 | 35.52 | 9.47 | 2.63 |
| 189 | 0.31 | 47.87 | 42.15 | 17.28 | 5.34 |
| 200 | 0.40 | 47.47 | 40.84 | 14.95 | 4.48 |

TABLE 1
SNAPTSHOT OF PRIVACY SCORES FROM FIGURE 6

$T^0$, $P_n$ and $C_n$, privacy score can be computed as in equation 4 below:

$$Naive\_Privacy\_Score = (\frac{P_n}{t_x - 1}) * (1 - C_n) * (T^0) \quad (4)$$

Figure 7 shows privacy scores derived from privacy scoring function $PS(X_n)$ against the Naive approach discussed above. Change in ego users disposition $l^0$ and $h^0$, non-linearly changes the privacy scores of friends while Naive scores are near constants in social network setting.

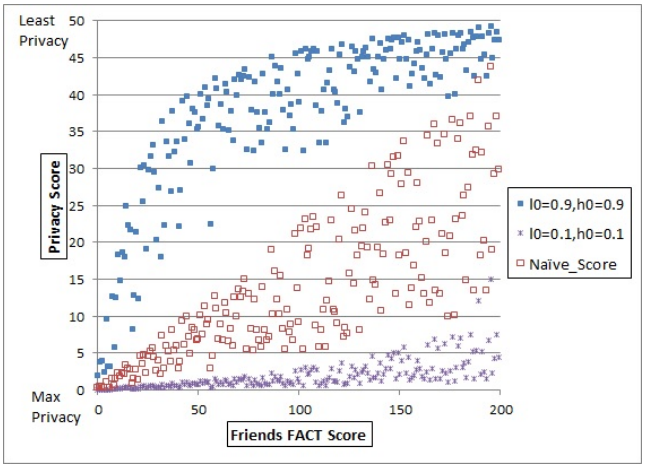


Fig. 7. Naive vs Proposed privacy scoring

## VIII. CONCLUSION AND FUTURE WORK

We have proposed a privacy scoring framework for calculating privacy scores of friends who are part of an ego-network. We have utilised ego users' disposition to privacy and communication in calculating privacy scores. Adapting cubic bezier curve in privacy score function, privacy score (ordinate) is determined as a function of FACT function (abscissa). Privacy scores calculated using ego users' disposition to privacy and communication paves way for communication based on categories, with each category housing for example,

sensitive, medium sensitive or non-sensitive information and access (to friends) to it controlled through privacy scores. Higher the sensitivity of information being disclosed, higher privacy orientation (lower Privacy score) that is required for a friend to receive that information.

In the current setting, where users are forced to group friends on each social network separately and having to comprehend diverse and non-overlapping privacy controls, Privacy scoring as a service is a better alternative. Since disposition of an ego user is non-changing with different social networks, privacy scoring service could serve the ego user across all social networks.

As our future work, we would like to test the proposed framework on real social network data. We believe the results to be similar to the experimental results discussed. We would also like to improve the FACT function by enhancing the friends sorting technique further.

## REFERENCES


[1] A. Acquisti and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. In *Privacy enhancing technologies*, pages 36–58. Springer, 2006.

[2] I. Ajzen and M. Fishbein. Understanding attitudes and predicting social behaviour. 1980.

[3] S. Amershi, J. Fogarty, and D. Weld. Regroup: Interactive machine learning for on-demand group creation in social networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 21–30, 2012.

[4] N. B. Ellison et al. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230, 2007.

[5] N. B. Ellison, J. Vitak, C. Steinfield, R. Gray, and C. Lampe. Negotiating privacy concerns and social capital needs in a social media environment. In *Privacy online*, pages 19–32. Springer, 2011.

[6] M. Eslami, A. Aleyasen, R. Z. Moghaddam, and K. Karahalios. Friend grouping algorithms for online social networks: Preference, bias, and implications. In *Social Informatics*, pages 34–49. Springer, 2014.

[7] M. Fishbein and I. Ajzen. *Belief, attitude, intention and behavior: An introduction to theory and research*. 1975.

[8] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010.

[9] J. Grimmelmann. Saving facebook. 2008.

[10] R. Gross and A. Acquisti. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80, 2005.

[11] W. Itani, A. Kayssi, and A. Chehab. Privacy as a service: Privacy-aware data storage and processing in cloud computing architectures. In *Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*, pages 711–716, 2009.

[12] J. J. Jones, J. E. Settle, R. M. Bond, C. J. Fariss, C. Marlow, and J. H. Fowler. Inferring tie strength from online directed behavior. *PloS one*, 8(1):e52168, 2013.

[13] S. Jones and E. O'Neill. 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, 2010.

[14] I. Kahanda and J. Neville. Using transactional information to predict link strength in online social networks. In *ICWSM*, 2009.

[15] S. Kairam, M. Brzozowski, D. Huffaker, and E. Chi. Talking in circles: selective sharing in google+. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1065–1074. ACM, 2012.

[16] P. G. Kelley, R. Brewer, Y. Mayer, L. F. Cranor, and N. Sadeh. An investigation into facebook friend grouping. In *Human-Computer Interaction–INTERACT 2011*, pages 216–233. Springer, 2011.

[17] K. Liu and E. Terzi. A framework for computing the privacy scores of users in online social networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 5(1):6, 2010.

[18] E. M. Maximilien, T. Grandison, T. Sun, D. Richardson, S. Guo, and K. Liu. Privacy-as-a-service: Models, algorithms, and results on the facebook platform. In *Proceedings of Web*, volume 2, 2009.

[19] R. K. Nepali and Y. Wang. Sonet: A social network model for privacy monitoring and ranking. In *IEEE 33rd International Conference on Distributed Computing Systems Workshops (ICDCSW)*, pages 162–166, 2013.

[20] A. Noack. Modularity clustering is force-directed layout. *Physical Review E*, 79(2):026102, 2009.

[21] K. Raynes-Goldie. Aliases, creeping, and wall cleaning: Understanding privacy in the age of facebook. *First Monday*, 15(1), 2010.

[22] R. Saadi, J.-M. Pierson, and L. Brunie. T2d: A peer to peer trust management system based on disposition to trust. In *Proceedings of the 2010 ACM Symposium on Applied Computing*, pages 1472–1478, 2010.

[23] M. Sramka. Evaluating privacy risks in social networks from the users perspective. In *Advanced Research in Data Privacy*, pages 251–267. Springer, 2015.

[24] A. Srivastava and G. Geethakumari. Measuring privacy leaks in online social networks. In *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 2095–2100. IEEE, 2013.

[25] K. Subrahmanyam, S. M. Reich, N. Waechter, and G. Espinoza. Online and offline social networks: Use of social networking sites by emerging adults. *Journal of Applied Developmental Psychology*, 29(6):420–433, 2008.

[26] Z. Tufekci. Can you see me now? audience and disclosure regulation in online social network sites. *Bulletin of Science, Technology & Society*, 28(1):20–36, 2008.

[27] B. Vidyalakshmi, R. K. Wong, M. Ghanavati, and C. H. Chi. Privacy as a service in social network communications. In *IEEE International Conference on Services Computing (SCC)*, pages 456–463, 2014.

[28] D. Wang, X. Liu, and X. Li. Blind spots: Unveiling users' true willingness in online social networks. In *IEEE Global Communications Conference (GLOBECOM)*, pages 2066–2071, 2012.

[29] Y. Wang, R. K. Nepali, and J. Nikolai. Social network privacy measurement and simulation. In *International Conference on Computing, Networking and Communications (ICNC)*, pages 802–806. IEEE, 2014.

[30] A. F. Westin et al. The dimensions of privacy: A national opinion research survey of attitudes toward privacy. 1979.

[31] A. F. Westin, D. Maurici, L. Price Waterhouse, and L. Harris. *E-commerce & privacy: What Net users want*. Privacy & American Business Hackensack, NJ, 1998.

[32] R. Xiang, J. Neville, and M. Rogati. Modeling relationship strength in online social networks. In *Proceedings of the 19th international conference on World wide web*, pages 981–990. ACM, 2010.