# User Identities across Social Networks: Quantifying Linkability and Nudging Users to Control Linkability

**Student Name: Srishti Chandok**
IIIT-D-MTech-CS-IS-17-MT15061
July, 2017

Indraprastha Institute of Information Technology
New Delhi

<u>Thesis Committee</u>
Dr. Ponnurangam Kumaraguru (Chair)
Dr. Arun Balaji Buduru, IIIT-Delhi
Dr. Anuja Arora, JIIT

Submitted in partial fulfillment of the requirements
for the Degree of M.Tech. in Computer Science,
in Information Security Category

# Certificate

This is to certify that the thesis titled **"User Identities across Social Networks: Quantifying Linkability and Nudging Users to control Linkability"** submitted by **Srishti Chandok** for the partial fulfillment of the requirements for the degree of Master of Technology in Computer Science & Engineering is a record of the bonafide work carried out by her under our guidance and supervision in the PreCog Research group at Indraprastha Institute of Information Technology, Delhi. This work has not been submitted anywhere else for the reward of any other degree.

**Dr. Ponnurangam Kumarguru**

**Indraprastha Institute of Information Technology, New Delhi**

# Abstract

The Online Social Network (OSN) landscape has transformed significantly over the past few years with the emergence of networks. The primary capabilities of these online networks differ. Few of the major leading ones are: Relationship networks (Facebook), Media sharing networks (Instagram), Online reviews (Zomato), Discussion forums (Quora), Social publishing platforms (Twitter), etc. In order to avail these services, users end up creating multiple identities across these platforms. For each OSN, a user defines his identity with a different set of attributes, genre of content and friends to suit the purpose of using that OSN. Researchers have proposed numerous techniques to resolve multiple such identities of a user across different platforms. However, the ability to link different identities poses a threat to the users privacy; users may or may not want their identities to be linkable across networks. In this study, we model the notion of linkability as the probability of an adversary (who is part of the users network) being able to link two profiles across different platforms, to the same real user. The major factors that lead to increased linkability across social networks are similar profile attributes and cross posting across the social networks. To make users aware of the linkability across multiple social networks, as part of the thesis, we develop a framework, which assists the users to control their linkability. It has two components; a linkability calculator that uses three state-of-the-art identity resolution techniques to compute a normalized linkability measure for each pair of social network platforms used by a user, and a soft paternalistic nudge. The user configures the desired linkability score range for each pair of networks. There are two types of nudge: *Attribute-driven Notification Nudge*, which alerts the user through a pop-up notification if any of their activity violates their preferred linkability score range and *Content-driven Color Nudge*, which notifies the user by changing the color of the box bounding the post update from black to red if the content being posted by them is found to be similar to the content already posted

by them on a different social network. We evaluate the effectiveness of the nudge by conducting a controlled user study on privacy conscious users who maintain their accounts on Facebook, Twitter, and Instagram. Outcomes of user study confirmed that the proposed framework of providing nudge helped 75% of participants to take informed decisions, thereby preventing inadvertent exposure of their personal information across social network services. Also, the content driven color nudge refrained few participants from making post updates.

# Acknowledgements

I would like to express my deepest gratitude to my advisor Dr. Ponnurangam Kumaraguru, "PK" for his guidance and support. The quality of this work would not have been nearly as high without his well-appreciated advice. I thank all the members of Precog research group at IIIT-Delhi for their valuable feedback and suggestions, especially Rishabh Kaushal for shepherding me and spending his valuable time to come up with this thesis. Last but not the least, I would like to thank all my supportive family who encouraged and kept me motivated throughout the project.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Research Motivation and Background

## 1.1 Research Motivation

Online Social Networks (OSNs) are becoming popular among users of Internet. These networks provide different types of services ranging from personal networks to interest based networks [4]. With so many social networks around, there are many reasons for users to register and maintain accounts (identities) across more than one OSN. [1] A few of the reasons are a)*type of content* being shared and b)*type of network* being offered. Examples of varying type of content is that some OSNs promote sharing of images (like Flickr and Instagram) or videos (like YouTube) while others promote sharing of short messages (like Twitter) or combination of messages, video and images (like Facebook). To offer different types of networks being provided to users, some OSNs provide access to professional network (like LinkedIn) while others provide access to a more personal network (like Facebook). These factors affect users' participation in these networks. For instance, an incoming friend request on a professional network tends to be accepted even if a requester is not personally known (referred as 'others') whereas on a personal network, a user would not like to accept such a request. Similarly, a user is likely to post about personal life events on a network like Facebook, but would probably refrain from doing the same on a professional network like LinkedIn.

There have been a number of instances in the past where social networks users were penalized for sharing content with probably the wrong audience. In 2009, Connor Riley, a master student at University of California, Berkeley, lost her job at Cisco before she could actually join it. She had tweeted about her job and did not bother to change the privacy setting that would have made the message exclusive to her friends. Her tweet received a reply from someone who claimed to pass her words to the hiring manager.

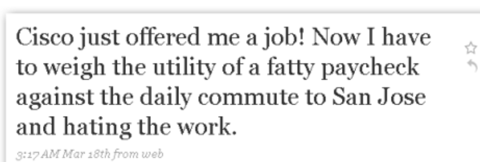



Figure 1.1: Tweet by Connor Riley.

A Georgia high school teacher, Ashely Payne had some students as her Facebook friends. When on vacation, she posted a picture with a glass of wine in one hand and beer in another. Upon discovering this content, one of the student's parent anonymously made the school authorities aware of it, which led Ashley loose her job.

[1] According to statistics released by Pew Research Center in 2016, more than half of online users (56%) use more than one OSN, a trend which has been consistent in the past few years [5].


Figure 1.2: A teacher fired for posting images of her vacation.

Linking of identities across online social networks have had severe consequences in the past. For instance, in 2012, a man was found spreading false information about the Superstorm Sandy from his Twitter account. Although the twitter account had no real name, Buzzfeed contributor Jack Stuef was able to identify the man by comparing photos posted on the account to unedited versions of them on a Tumblr account. The account owner was identified as Shashank Tripathi, a hedge fund analyst and campaign manager for a candidate of a political party. Due to such severe

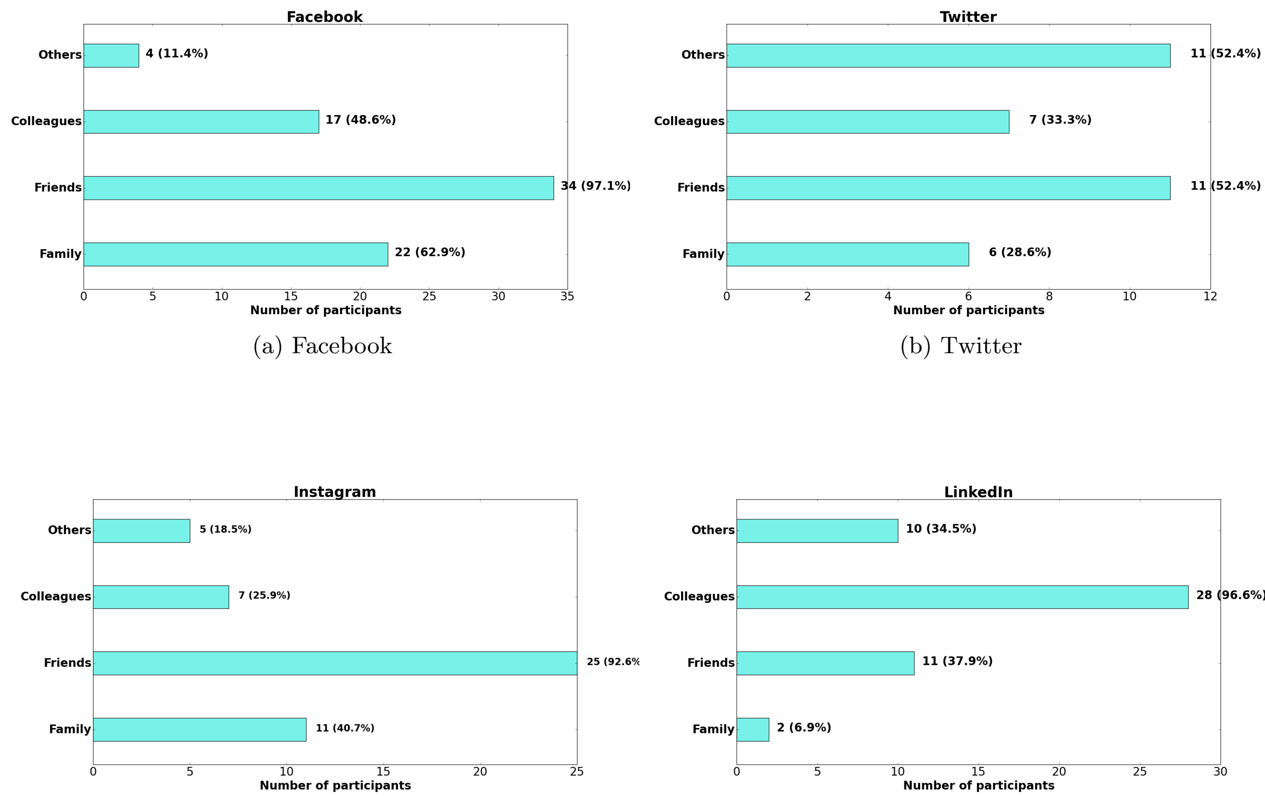

Figure 1.3: Resolving of identities got a member replaced from political affiliation.

repercussions (as discussed in previous section) of posting content on social network, there arises a need for users to become more aware of identity linkage. A potential solution could be to build a system that tells how much linkable a user's social networks are and then to further nudge him whenever his actions lead to increased linkability across the networks.
Current research has explored Identity Resolution [7, 8, 9, 10, 11, 12, 13, 14, 16, 18]. However, to the best of our knowledge less attention has been paid on (a) Quantifying linkability in form of a metric called *Linkability Score*, (b) Linkability Nudge, system to help users control their linkability and provide soft interventions whenever a user behavior causing linkability to go beyond desired range. In this work we take a step ahead to fill this gap.

## 1.2 Background

We conducted an online survey based on a *pre-study questionnaire* to gain insights to users behavior across online social networks, refer Appendix for "Questions". Through the survey, we validate that users register at multiple OSNs for different purposes. It is evident that family and friends dominate participants' connections on Facebook & Instagram while on Twitter, more than

50% of respondents say that 'others' (non-acquaintance) also form a predominant chunk of audience and on LinkedIn, around 96% of respondents say that they are connected to their colleagues, refer Fig 1.4.




(a) Facebook

(b) Twitter

(c) Instagram

(d) LinkedIn

Figure 1.4: Participant responses about the intended audience on Facebook, Twitter, Instagram and LinkedIn.

Another observation made is regarding the content which users tend to post on OSNs. Users post similar content (cross-posting) on some pair of OSNs while in other pairs, they don't cross-post. More than 75% of respondents said that they do post similar content on Facebook - Instagram as depicted in Fig 1.5. Posting similar content across OSNs increases the chances of a person being linked across those OSNs.

We further asked respondents, of their comfort level if a user $X$ who is their friend (or connected) on OSN $A$ can identify them or their activities on another OSN $B$ where $X$ is not their friend. It was found that when $A$ = Twitter and $B$ = Facebook, Instagram or LinkedIn, 40-50% of participants were not comfortable, as depicted in Fig 1.6. A plausible explanation of this could be that on Twitter a number of non-acquaintances ('others') also become followers and it would **not** be desirable if they get access to user's activities on other OSNs like Facebook and Instagram which have user's family relatives and friends.

Through these observations it is clear that on one hand there are some pairs of OSNs where users have no issues in maintaining their identities as close to each other while on other hand, on other pairs of OSNs, they tend to keep their identities as far as possible. We capture this notion using a term called *linkability* of user identities across OSNs. However, similar identities across OSNs give rise to a variety of privacy implications which are seldom addressed or acknowledged.

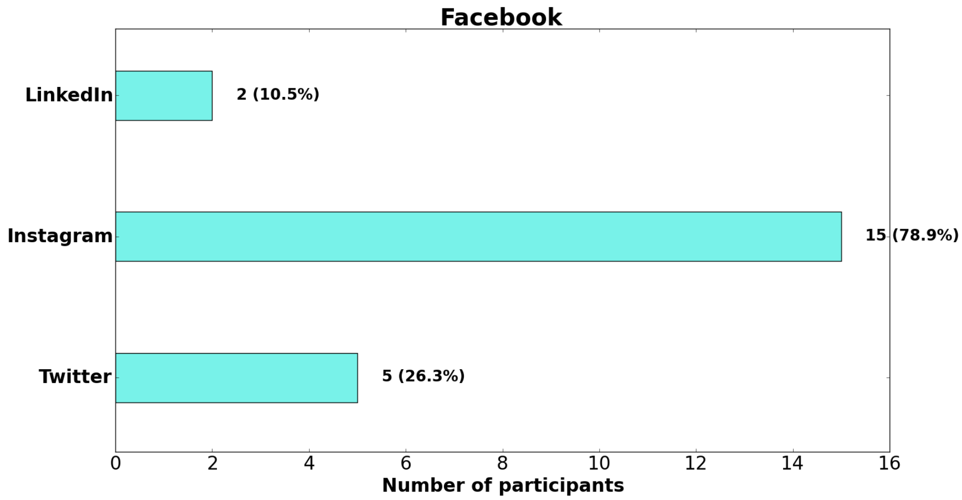



Figure 1.5: Participant responses about cross-posting (post same content) content from Facebook to other social networks.

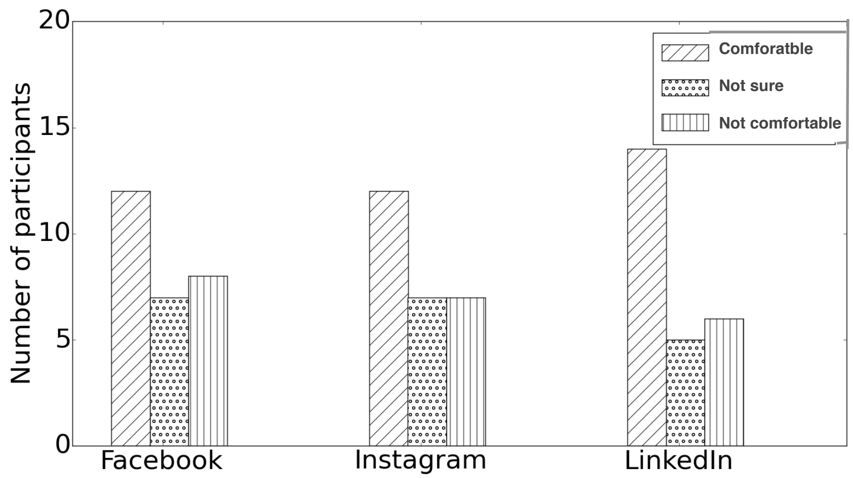
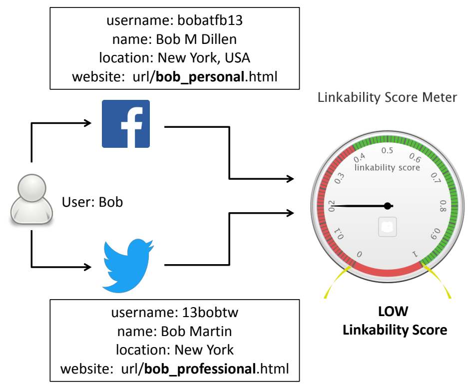


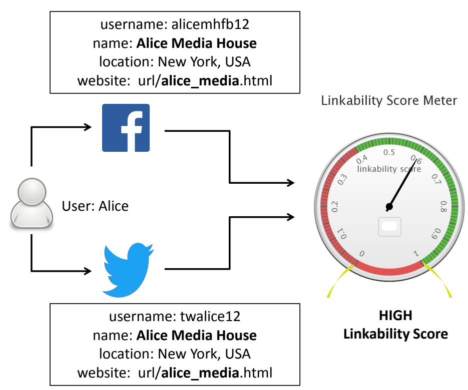

Figure 1.6: Participant responses about a user 'X' who is your friend on Twitter, gets to see your activities on Facebook, Instagram and LinkedIn, what would be your comfort level?

Consciously or unconsciously, users tend to have a certain set of attributes and characteristics common across multiple social networks (for example, date of birth, city of residence, screen name, etc.), which enables third parties and adversaries to be able to link two profiles on different platforms to the same real world user. While some users may not care about their profiles being linked and others might, in most cases, users are simply unaware of the phenomenon of any such *linkability* of their profiles. To illustrate this further, in Fig 1.7a, we depict a user named **'Bob'** who consciously maintains his two identities to keep his personal and professional network separate on two OSNs namely Facebook and Twitter. He does so by keeping various attributes like username, name and URL different on both OSNs. As a result, his linkability score is 0.2 which is quite less, thereby indicating low linkability.

In a complementary situation, a user who wants to publicize a brand would create identities across different OSNs so as to reach to larger audiences. Contrary to earlier scenario, in this case




(a) User Bob maintaining separate professional and personal identities across more than one OSN (Twitter and Facebook), hence keeping low linkability score.




(b) User Alice maintaining multiple identities across more than one OSN (Twitter and Facebook) for branding a Media Company desirous of keeping high linkability score.

Figure 1.7: Two complementary user scenarios with respect to linkability.

such a user would not mind *linkability* of identities but rather would like that their identities to be as similar (consistent) as possible so as to avoid any confusion among their audience about the brand (refer Fig 1.7b). Therefore, user may be interested to know *linkability* of identities so as to keep it high in order to enable audiences to reach to these multiple identities. Although this is not the scenario we target from privacy point of view, we have still mentioned it here to draw home the point that linkability score would be useful in such a scenario as well.

Researchers have termed this concept of linking two online profiles to a user as *identity resolution*, and have demonstrated multiple techniques in the past where they have been able to correctly link profiles across platforms with a high success rate [7, 8, 9, 10, 11, 12, 13, 14, 16, 18]. Looking from within the research community, we find that there exist a problem called *identity resolution*, in which the goal is to find (or *link*) all identities belonging to the same user across multiple OSNs. Numerous methods exist for performing identity resolution. One of the key motivations of solving identity resolution problem is to build a comprehensive profile of a user to be used for better online marketing. However, on flip side, the same methods can be exploited by an attacker to link user's identity on professional network to that on personal network, thereby gaining access to personal information. Consequently, a privacy conscious user (whom we focus in this thesis, also evident from responses of *pre-study questionnaire*, in which majority of the participants said that they were quite aware of privacy settings of OSNs which they use, refer Fig 1.8) would like to keep his identity on a personal network as separate as possible (*non-linkable*) from that on a professional network.

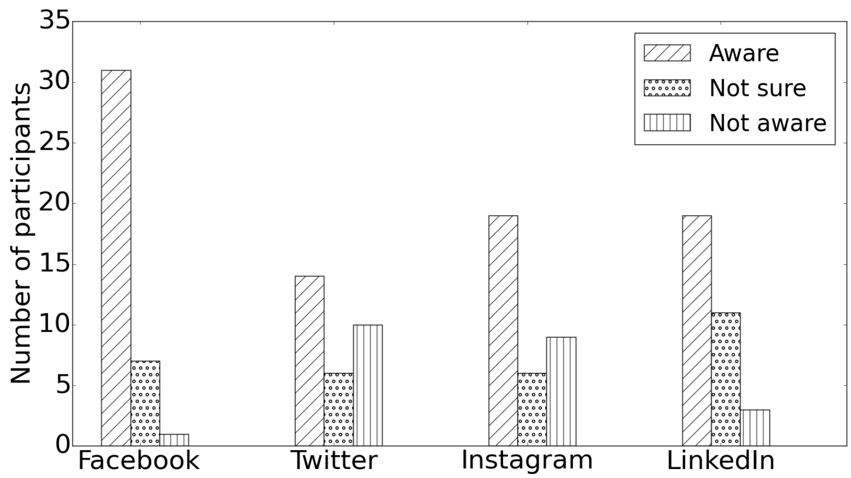



Figure 1.8: Participant responses about awareness of the privacy settings or controls in following social network.

## 1.3 Organization of Thesis

The rest of the thesis is organized as follows. The research aim and various contributions of the thesis are described in Chapter 2. Chapter 3 describes related work to methods for Identity Resolution and privacy nudges. Subsequently, in Chapter 4, we discuss about the baseline approach for computing linkability scores. In Chapter 5, we explain the reformed linkability score computation by leveraging well known Identity Resolution methods and discuss about the design of linkability nudge. In Chapter 6, we explain in detail architecture of proposed linkability nudge, followed by user evaluation of nudge and results. Finally in Chapter 7, we discuss about the implications and limitations of our work and outline the future work.

# Chapter 2

# Research Aim and Contributions

## 2.1 Research Aim

We aim to demonstrate user's linkability across various social networks and to develop a real-time system which can help users to maintain their linkability across social networks. We intent to closely monitor and analyze the user actions on OSNs and accordingly nudge the user when the linkability range crosses the pre-configured range.

## 2.2 Contributions

In our study, we address the privacy implications introduced by identity resolution, and propose Nudging Nemo, a framework that allows users to learn about, and control the linkability of their own profiles across different social networks. Our key contributions are as follows:

- We propose a *linkability score* metric which quantifies separation / closeness between two identities belonging to the same user on different OSNs and suggest baseline methods based on weighted sum and probabilistic to compute the metric.

- Evaluate the efficacy of proposed metric on a large pair of user identities drawn from Facebook and Twitter. To be specific, we evaluated our baseline methods on 23,985 true positive identity pairs.

- Compute the linkability scores by implementing well known methods for identity resolution (like NEMO [10], HYDRA [18] and MOBIUS [7]). Along with the linkability score computation, we also identify the factors (profile attributes) that have contributed to the computed linkability score. This ensures that user is better informed to undertake corrective measures so as to bring down (or up as the case may be) the linkability score between a pair of identities on two social networks.

- Design and develop linkability nudge, a soft paternalistic intervention [19, 20, 21, 22, 23, 24, 25] which alerts users whenever user behavior leads to significant change in linkability score beyond preconfigured range. The nudge is designed to work in two stages. In *first stage*, user preferences are taken by allowing users to configure a range of desired linkability score for each pair of OSN platform. User is also provided an option to *opt-out* by way of which user won't be subjected to nudges. In *second stage*, user's behavior in terms of content being posted or changes in profile attributes are monitored. Users are alerted (nudged) whenever their behavior changes their linkability score beyond a desired range. Nudges are rendered in two forms, one through *colors* and other through a *pop-up notification message*, explained in detail subsequently.

- Perform a detailed user study in a controlled lab experiment setting to assess effectiveness and utility of proposed linkability nudge.

# Chapter 3

# Related Work

Work related to our problem can be divided into two broad categories. In *first category*, we look at research in the area of *identity resolution*, which is the conventional term used in research literature for finding multiple identities of the same user across different OSNs. The goal in these works is to propose novel methods to search and link identities belonging to the same user across OSNs. Identity resolution approaches for linking identities are leveraged in computation of linkability score in this thesis. In *second category*, we explore works done wherein users are suggested (or *nudged*) to change their OSN usage and behavior with the goal of reducing privacy disclosures due to certain factors, we refer these methods collectively as *privacy nudges*. Ideas taken from privacy nudges are used in developing *linkability nudge* in our work.

## 3.1 Identity Resolution

Numerous methods and techniques have been studied by researchers for performing identity resolution across multiple OSNs. Work from Zafarani et. al. [7] exploited users' unique behavioral patterns that lead to information redundancies across sites to solve the problem of identity resolution. Bartunov et. al. [8] proposed an approach based on conditional random fields which leverages profile attributes and social linkages. As claimed, the approach performs better than attribute-based approaches which rely only on profile attributes. Carmagnola et. al. [9] proposed an approach to identify users on different social networks using their public data. Jain et. al. [10], for instance, divided the task of identity resolution into two steps namely, *identity search* and *identity match*. Identity search takes a user's identity on OSN platform A and searches for candidate identities on other OSN platforms, while, identity match then links the most closest among the candidate identities to the given input user's identity. Identity search was performed based on features taken from profile, content, self-mention and network. For identity match, syntactic matching (Jaro distance) was used for textual features and RGB-histogram based image matching was used to match profile images. The author presented novel methods [11] for searching and linking multiple identities across OSNs. Their eventual aim was to aggregate the data related to same user across multiple sites by identifying users and retrieving their data. According to Liu et. al. [12], the task of linking users across communities, comprises of two variants. In first variation, task is to find whether a set of usernames are owned by a single person and in second, the task is to find all usernames of a single person across multiple sites. Goga et. al. [13] explored the innocuous activities of users namely in the form of geo-location attached to user's posts, timestamp of posts and writing styles of users to identify accounts that belong to the same user across different sites. Lim et. al. [14] studied the cross-sharing behavior (which refers to the act of posting similar content across multiple sites) of users across six OSNs namely Flickr, Google+, Instagram, Tumblr, Twitter and YouTube. Zhang et. al. [15] proposed canopying framework that performs linking of user profiles by using domain knowledge from online social networks. Li et. al. [16] uses both personal and social identity features of users to develop better identity matching techniques. All this literature on identity resolution focuses on identifying novel approaches and features to search for potentially matching identities and then resolving those identities using the best possible match. However, in this thesis, we look at the issue from user's perspective and propose a quantifiable measure for linkability of two identities, which we refer as *linkability score*. More specifically, we draw inspirations from three well know identity resolution methods, referred henceforth as NEMO [10], HYDRA [18] and MOBIUS [7] in computation of linkability scores as shall be explained later.

## 3.2 Privacy Nudges

In this section, we refer the work related to developing methods for informing and suggesting users with the aim of preventing privacy disclosures. In this thesis, we have taken ideas from these works to develop our proposed system that informs users about linkability score for each pair of OSNs and nudges them whenever their behavior causes score to go beyond desired range. Privacy nudge can be defined as a mechanism of introducing a suggestion to the users with the purpose of cautioning them against privacy disclosures which they might overlook. Typically the design of privacy nudge mechanisms would involve making changes in the user interface taking care that it ought not cause discomfort to the user and yet at the same point provide user with a useful suggestion safeguarding the user from privacy disclosures. Leenes et. al. [17] in their work suggested segregation of audience for profile attributes of user on OSNs so that its visibility is controllable. Wang et. al. [19] designed and implemented modifications to the Facebook web interface that would nudge user to consider the content and audience of their online disclosures. Wang et. al. [20] had also earlier developed three types of privacy nudge, one was to provide the audience of a post, second was developed to introduce time delays before a post goes public and third was provided to obtain user feedback. Authors in [21] and [22] worked to understand and find out the set of actions that users perform over OSNs which they later regret which could be a good indicator of privacy leaks and need for nudging so that those actions don't get repeated in future. Ziegeldorf et. al. [23] proposed a novel design paradigm called *comparison based privacy* in which a users can compare their privacy metrics with other groups of users to evaluate privacy disclosure levels. From works of [24] and [25], it can be seen that with widespread use of mobile devices, the ideas of privacy nudges are being applied on mobile platforms as well. To the best of our knowledge, there exist no prior work which provides a mechanism of nudging (or providing a feedback) users to prevent disclosures owing to the resolution of their multiple identities, which is the focus of this thesis.

# Chapter 4

# Linkability Score: Baseline Computations and Evaluation

In this chapter, we discuss our baseline computations for linkability score, understand its distribution across all types of user pairs and evaluate the efficacy of linkability score in estimating the extent that two given identities are linkable. We compute linkability scores taking into consideration the two most popular social networks, Facebook and Twitter.

## 4.1 Threat Scenario

We presume that attacker would have access to victim's identity on at least one OSN (say $i_A$) and would subsequently use one or more of the multiple variations of identity resolution as below.

(1) Given a pair of identities $i_A$ and $i_B$ on two OSNs $A$ and $B$, respectively, the goal is to find a function that returns 1 or 0 depending upon whether $i_A$ and $i_B$ belong to same user or not, respectively.

(2) Given a single identity $i_A$ on OSN $A$ and candidate set of identities $C_B$ on OSN $B$, the goal is to find a function which identifies correctly $i_B$ from within $C_B$ (searching problem).

(3) Given matching set of identities $C_A$ and $C_B$ on two OSNs $A$ and $B$, respectively, the goal is to find all pairs of identities ($i_A$ , $i_B$) which belong to the same user.

Attacker would typically implement well known methods that solve identity resolution problem taking $i_A$ as input and obtain $i_B$ of victim in OSN $B$ in which the adversary is not connected to the victim. In the context of earlier discussions this would mean that an adversary could be friend of victim in a professional network and using identity resolution methods, adversary could identify victim in a personal network, thereby gaining access to victim's activities in $OSN_B$. Given that attacker would leverage well known IR methods, we proceed ahead in Chapter 6 with their implementations. We also propose a metric in this chapter, which we call as *linkability score* that indicates possibility of success for a potential attacker.

## 4.2 Proposed Approach

In this section, we propose our methods for calculation of *linkability score* between two given identities $< I_a, I_b >$. The two identities may be drawn from either same or separate OSN, it won't affect computation of linkability score. We use the notation of $f^i$ to denote the $i^{th}$ *feature* (username, name of user, location, etc.) and use the notation of $m_j$ to denote the value of $j^{th}$ *metric* (hamming distance, cosine distance, longest common subsequence, etc.) computed between $i^{th}$ *identities*. $F$ and $M$ represent total number of features and metrics being considered, respectively.

### 4.2.1 Weighted Sum Method

Given an identity pair $< I_a, I_b >$ as input, the linkability score according to weighted sum method is a function of features ($F$) and metrics ($M$) used to calculate similarity between features.

$$L_S \leftarrow f_{weighted-sum}(F_M) \tag{4.1}$$

Computation of linkability score is performed in two steps namely *feature similarity indicator* and *linkability score calculator*.The intuitive idea here is the linkability score depends additively in different proportion on various features and similarity metrics.

In *feature similarity indicator (FSI)* step, we find similarity between $i^{th}$ feature corresponding to the two identities $I_a$ and $I_b$ as below.

$$FSI(f_a^i, f_b^i) = \frac{1}{M} \cdot \sum_{j=1}^{M} m_j(f_a^i, f_b^i) \tag{4.2}$$

In *linkability score calculator* step, we eventually find the linkability score taking into account all the features. Higher is the linkability score, more is the possibility of identities getting resolved through automated methods and lesser is cross-OSN anonymity.

$$L_S(I_a, I_b) = \frac{1}{\sum_{i=1}^{F} wt_i} \cdot \sum_{i=1}^{F} wt_i \times FSI(f_a^i, f_b^i) \tag{4.3}$$

In eq (5.2), $wt_i$ represent the weights assigned to the feature. Higher the weight, more is the contribution of the feature in the linkability score computation. We describe a method for estimating these weights later in this section. Algorithm 1 explains the computation of linkability scores using weighted sum method. It takes as input all identity pairs collected in dataset, denoted by $D$.

The inner loop at *Steps* 6 in Algorithm 1 finds the feature similarity indicator, whereas the outer loop at *steps* 5 calculates linkability score of $k^{th}$ *identity pair* $< I_a, I_b >$. *Steps* $3 - 12$ execute for all identity pairs in the dataset $D$ and finds linkability scores for all identities in it. Running time for the algorithm is $\theta(nFM\sigma)$, where $n$ is number of identity pairs in dataset $D$ and $\sigma$ is the cost of computing *cal_metric*() function.

**Feature_Weight_Estimation**: Feature weights $w_i$ associated with eq (5.2) for each of the $i^{th}$ feature are **estimated** based on linkability scores calculated for positive and negative data. Higher is the value of weight, more prominent contribution is made by the feature in the linkability score. For estimating feature weights, we follow an approach for *maximization* that has to ensure that difference between linkability scores for negative data (Type I) ($ND_1$) & negative data (Type II) ($ND_2$), represented by eq (4.4) and also the difference between linkability scores for positive data ($PD$) & negative data (Type I) ($ND_1$), represented by eq (4.5), is maximized.

$$diff_{ND_1-ND_2} = \sum_{i \in ND_1} L_S(i) - \sum_{j \in ND_2} L_S(j) \tag{4.4}$$

**Algorithm 1** To calculate Linkability Score using Weighted Sum Method

**procedure** LS_WEIGHTED_SUM($D$)
  $LS \leftarrow \emptyset$
  **for all** $k^{th} identity\_pair < I_a, I_b > \in D$ **do**
    $LS_k \leftarrow 0$
    **for all** $i \in F$ **do**
      **for all** $j \in M$ **do**
        $FSI_i \leftarrow \frac{\sum_j cal\_metric_j(f_a^i, f_b^i)}{M}$
      **end for**
      $LS_k \leftarrow \frac{\sum_i wt_i \times FSI_i(f_a^i, f_b^i)}{wt_{sum}}$
    **end for**
    $LS \leftarrow LS \cup LS_k$
  **end for**
**return** $LS$
**end procedure**

and

$$diff_{PD-ND_1} = \sum_{i \in PD} L_S(i) - \sum_{j \in ND_1} L_S(j) \tag{4.5}$$

Algorithm 2 explains the procedure for estimating the weights of various features for computing linkability score.

The algorithm for weight estimation depends upon the function $get\_weight\_combinations()$ which

**Algorithm 2** To estimate feature weights in Weighted Sum Method

1: **procedure** ESTIMATE_FEATURE_WEIGHT($w$)
2:   $n \leftarrow len(I_a) \leftarrow len(I_b)$
3:   $max\_diff_{ND_1-ND_2} \leftarrow 0$
4:   $max\_diff_{PD-ND_1} \leftarrow 0$
5:   **for all** $wt_i \in get\_weight\_combinations()$ **do**
6:     $LS_{ND_1} \leftarrow$ LS_WEIGHTED_SUM($D_{ND_1}$)
7:     $LS_{PD} \leftarrow$ LS_WEIGHTED_SUM($D_{PD}$)
8:     $LS_{ND_2} \leftarrow$ LS_WEIGHTED_SUM($D_{ND_2}$)
9:     $diff_{ND_1-ND_2} \leftarrow LS_{ND_1} - LS_{ND_2}$
10:     $diff_{PD-ND_1} \leftarrow LS_{PD} - LS_{ND_1}$
11:     $maximize(diff_{ND_1-ND_2})$
12:     $maximize(diff_{PD-ND_1})$
13:   **end for**
14: **end procedure**

would return all possible integral weight values to the $F features$, so that the total number of combinations would be *F!*. *Steps* $6-8$ in *Estimate_Feature_Weight()* procedure, collectively are computed over entire dataset $D$ comprising of $n$ identity pairs while *steps* $9, 10$ compute values as defined in eq (4.4) and eq (4.5), respectively. Running time for the entire weight estimation algorithm would be $\mathcal{O}(nF!)$. This computation is costly, however, it may be noted that weight estimation would be done as a *pre-processing step* in *offline* mode and would be repeated only when there is substantive change in the dataset. Currently a *brute force* approach is used, however, as the number of features increase, we can employ an efficient algorithm for optimized estimation of feature weights.

Detailed results for all the possible combinations of weights and the values of maximum differentiation obtained are explained in the evaluation section.

### 4.2.2 Probabilistic Method

In our second method, we use *probabilistic approach* to compute the linkability scores between two identities $I_a$ and $I_b$. The intuitive idea here is that linkability score depends probabilistically on similarity values of features. According to this method, linkability score is defined as the probability of finding when the two identities are **same** ($I_a = I_b$) conditional on the *similarity* of their features $<f^1....f^F>$ as explained below, only $i^{th}$ feature considered.

$$L_S \leftarrow prob\left((I_a = I_b)|(f_a^i = f_b^i)\right) \tag{4.6}$$

which can be expressed as below

$$prob\left((I_a = I_b)|(f_a^i = f_b^i)\right) = \frac{prob\left((I_a = I_b) \cap (f_a^i = f_b^i)\right)}{prob(f_a^i = f_b^i)} \tag{4.7}$$

Here, $prob\left((I_a = I_b) \cap (f_a^i = f_b^i)\right)$ is scenario in which two identities belong to same user and feature similarity between $i^{th} feature$ of $I_a$ and $I_b$ is $X$ such that $f_a^i = f_b^i = X$. And, $prob(f_a^i = f_b^i)$ is probability that $i^{th} feature$ of $I_a$ and $I_b$ takes value as $X$ for all scenarios, namely, positive data and negative data (Type I and II). Algorithm 3 computes linkability score using probabilistic method. In it, *step 13* calculates $prob\left((I_a = I_b) \cap (f_a^i = f_b^i)\right)$ as $pd = \sum_{j=1}^{M} find\_prob(m_j, D_{PD})$. In order to find it, we do a look-up in the dataset of *positive data* ($D_{PD}$) and find the frequency value corresponding to $f_a^i = f_b^i = X$, so as to capture frequentist approach for calculating probabilities. This look-up is performed by function $find\_prob$ in Algorithm 3. Value of $prob(f_a^i = f_b^i)$ is computing by $pd_i + nd_{1i} + nd_{2i}$. Running time for the algorithm is $\theta(nFM\sigma)$, which is same as that obtained in weighted sum method.

## 4.3 Data Collection

We describe our approach for collection of ground truth data in this section which would eventually form input to the function that computes linkability score ($f_{linkability\_score}$). We focus on the scenario of identities belonging to different OSN sites ($a \neq b$). Therefore, data, for the purpose of our problem, comprises of pair of identities (tuple) $<I_a\ ,\ I_b>,\ where\ a = Facebook\ and\ b = Twitter$. With respect to owner of the identity pair, following two scenarios are possible.

(1) *Positive Data:* Identities in the pair $<I_{fb}\ ,\ I_{tw}>$ belong to the *same* user.

(2) *Negative Data:* Identities in the pair $<I_{fb}\ ,\ I_{tw}>$ belong to *separate* users.

We now discuss the approaches used for collecting these types of data.

### 4.3.1 Positive Data Collection

One of the biggest challenge is to collect identities of *same* user on two OSN sites. We have examined the users who share their content objects (posts, images, etc.) originally posted on other OSN sites like Facebook. This sharing of content appears in the tweets on Twitter in form

**Algorithm 3** To calculate Linkability Score using Probabilistic Method

**procedure** LS_PROBABILISTIC_METHOD($D$)
  $LS \leftarrow \emptyset$
  **for all** $k^{th} identity\_pair < I_a, I_b > \in D$ **do**
    $pd \leftarrow nd_1 \leftarrow nd_2 \leftarrow 0$
    **for all** $i \in F$ **do**
      $pd_i \leftarrow nd_{1i} \leftarrow nd_{2i} \leftarrow 0$
      **for all** $j \in M$ **do**
        $m_j \leftarrow cal\_metric_j(f_a^i, f_b^i)$
        $pd_i \leftarrow pd_i + find\_prob(m_j, D_{PD})$
        $nd_{1i} \leftarrow nd_{1i} + find\_prob(m_j, D_{ND1})$
        $nd_{2i} \leftarrow nd_{2i} + find\_prob(m_j, D_{ND2})$
      **end for**
      $pd \leftarrow pd + \frac{pd_i}{M}$
      $nd_1 \leftarrow nd_1 + \frac{nd_{1i}}{M}$
      $nd_2 \leftarrow nd_2 + \frac{nd_{2i}}{M}$
    **end for**
    $LS_{identity\_pair} \leftarrow \frac{pd}{(pd+nd_1+nd_2)}$
    $LS \leftarrow LS \cup LS_{identity\_pair}$
  **end for**
**return** $LS$
**end procedure**

of a shortened URL, referred as *'fb.me'* links. We leverage this *self-mention* user behavior and filter out all tweets containting 'fb.me' links, expand their URLs and identify the identity of this user on Facebook. For positive data collection, we begin by collecting Twitter data through its Streaming API.[2] Users' data on Twitter, by default, is available in public domain, so this data can be collected without any issues. Identity on Twitter can be found by looking at the user posting the tweet. Using this method, we found 23,985 identity pairs of same user on Twitter and Facebook as mentioned in Table 4.1. We subsequently collect the user's profile attributes corresponding to these identity pairs.

Table 4.1: Summary of Data Collected for all possible scenarios

| **Description of Data** | **Count** |
|---|---|
| Positive Data Instances | 23,985 |
| Negative Data Instances (Type I) | 96,130 |
| Negative Data Instance (Type II) | 24,560 |

### 4.3.2 Negative Data Collection

In order to introduce noise and bring our data as close to the real world, it is important that it contains identity pairs which belong to separate users, referred as *negative data*. There are two types of negative data that we collect. *Type I* negative data are those identity pairs which *appear* to be quite *similar*, however, the identities belong to different users. *Type II* negative data are those identity pairs which are quite *dissimilar* in appearance and belong to separate users.

[2] https://dev.twitter.com/streaming/overview

**Negative Data Type I:** To collect negative data of Type I, we use the identities on Facebook already collected for positive data as input. For each of these identities, we collect upto 4-5 identities on Twitter whose names are *similar* to names in the input identity. In total, we collect 96,130 negative data of Type I instances as mentioned in Table 4.1.

**Negative Data Type II:** To collect negative data of Type II, we pick random identities from Facebook and Twitter and pair them together. Since they are collected by random, they would appear to be quite dissimilar and would belong to separate users. We collect 24,560 negative data of Type II instances as mentioned in Table 4.1.

### 4.3.3 Features and Metrics

During the data collection phase, for each identity collected, we also collect information related to various profile attributes of the users, which we refer as *features*. We do this feature collection so as to assess similarity among profile attributes of identities of a user. Table 4.2 lists down the set of profile specific features comprising of username, full name of user, location of user and profile & background image of user. For each of these features, we have mentioned in Table 4.2, the suitable *similarity* metrics which would be a good measure to reflect similarity between two identities.

Table 4.2: Feature list along with their suitable metrics.

| Feature Name | Metrics |
|---|---|
| username | Hamming Distance, Longest Common Subsequence, Edit Distance Cosine Distance, Jaccard Distance, JaroWinkler Distance. |
| name of user | Length of Common Substring, Length of Common Prefix & Common Suffix, Q-gram Distance. |
| location | Length of Common Substring, Geo-location Latitude & Longitude. |
| image | Image similarity metrics. |

*Username* refers to the **unique** identifier within an OSN site to recognize a user. While creating identities across different OSNs, users try to keep their username as **similar** to their names as possible unless the username has already been taken by some other user. Metrics for evaluating similarity among userIDs like Hamming distance, Cosine distance, Longest Common Subsequence, etc. operate at **character level**.

*Name of user* refers to the first, middle and last name of a user. This is the most important user's feature that is typically used to **search** for an identity on an OSN site. However, at the same time, since more than one user can have same names, the search results returned often contain a number of false positive user profiles. Similarity metrics for user name are length of common substring, common prefix, common suffix, etc. which operate at **string level**.

*Location* refers to the geographical coordinates (latitude and longitude) of user of OSN site. Location alone doesn't appear to be very significant since there would be many users who may have same geo-location, however, when combined with other features (like username), it is very effective in narrowing down the candidate matches for identity resolution. Not every user configures (or makes it visible in public domain) location attribute, hence it is found in only a limited number of identities. The location could be same but can be referred with two or more different location names.

Therefore, values of latitude and longitude is the most important metric for finding similarity among locations.

*Image* refers to the profile image of a user available in public domain. Image gives a good estimate of the similarity of identities, however, metrics for finding image similarity are computationally more complex. We find ORB descriptors of profile images in user identities using openCV library. [3]

**Limitation:** It may be stated here that due to restrictions in the endpoints offered by APIs of Facebook and Twitter, only a limited number of profile features are available in public domain for programmatic access. Further, different OSNs have varying number of profile attributes (features), for instance, Facebook has a very large set of user's profile attributes like employment details, school details, etc. but most of them are inaccessible through APIs whereas Twitter don't have such explicit profile attributes but rather has a generic attribute of 'about me' where a user **can** mention his/her employment and/or school details. In this approach, we have not considered these features, however, linkability score computation can be extended to include them in future work.

## 4.4 Results

In this section, we present the results for computed linkability scores based on the proposed approaches of weighted sum method and probabilistic method discussed earlier.

### 4.4.1 Linkability Scores - Weighted Sum Method

For the purpose of calculating the linkability scores using the weighted sum method, we first start from the scenario in which all features have equal weights (i.e. $a_i = 1, \forall i \in F$) and compute the linkability scores using algorithm $Cal\_Wt\_Linkability\_Score()$ for all types of data as depicted in Fig 4.1. Values of $diff_{PD-ND1}$ and $diff_{ND1-ND2}$ (represented in eq (4.4) & eq (4.5) and computed in *steps* $9, 10$ of Algorithm 2) mentioned as 3950 and 4659 represent the summations of differences in linkability scores for identities belonging positive data & negative data (Type I) and negative data (Type I) & negative data (Type II), respectively.

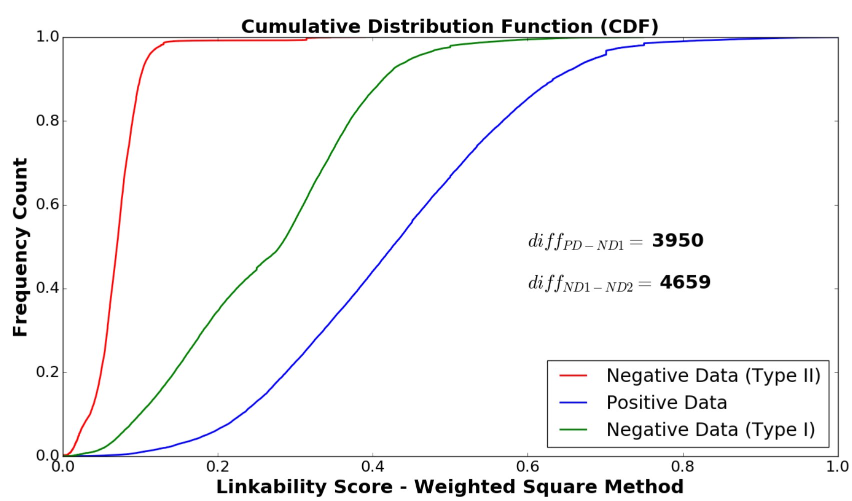



Figure 4.1: Comparison of linkability scores for all types of datasets assuming that all features have their weights equal to 1, Baseline Scenario.

[3] http://opencv-python-tutroals.readthedocs.io/

It can be observed from Fig 4.1 that linkability scores for 95% identities in negative data (Type II) is less than 0.1 which is as desired. Further it is evident that the distribution of linkability scores of identities belonging to negative data (Type I) is *less* than positive data scenarios. However, at the same time, we do find that overlap among their linkability scores within the range of 0.2 to 0.5 (also quantified as $diff_{PD-ND1} = 3950$) is high, which needs to be improved further.

Consequently, in the next analysis, we take into consideration the fact that features contribute with varying weights and perform the procedure of weight estimation outlined in algorithm $Feature\_Weight\_Estimation()$ with an aim to increasing $diff_{TP-FP}$ or in other words increasing the gap between distribution of linkability scores of positive data and negative data. In Fig 4.2, we draw a scatter plot of difference between linkability scores for negative data (Type I) & negative data (Type II) ($diff_{ND1-ND2}$) on Y-axis vs difference between linkability scores for positive data & negative data (Type I) ($diff_{PD-ND1}$).

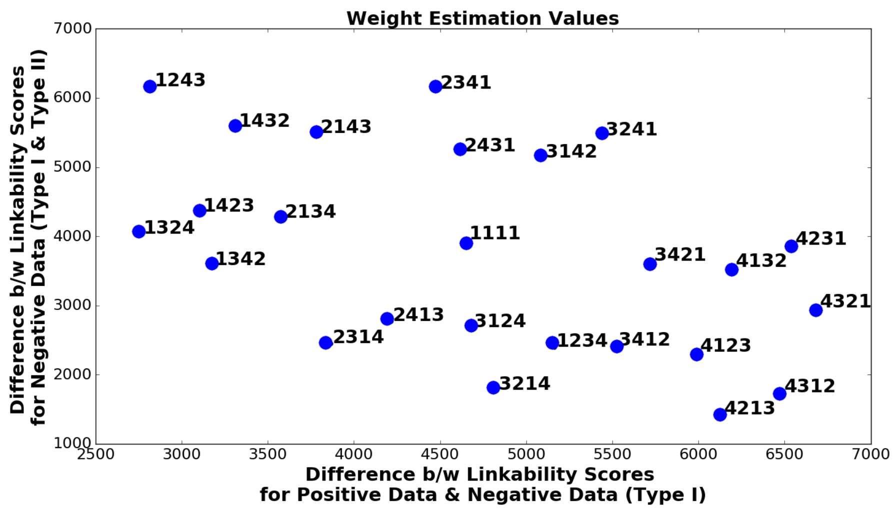



Figure 4.2: Linkability scores of $diff_{PD-ND1}$ vs $diff_{ND1-ND2}$, the four-digit values mentioned for each data point represent the weights assigned to four features considered namely username, name, geo-location and website/url, respectively. Fifth feature of image has been skipped so that combinations of weights remain limited and data points can be easily visualized.

Upper left quarter comprises of those weight combinations in which difference between linkability scores for positive data and negative data (Type I) ($diff_{PD-ND1}$) is high, for instance weight combinations 1243, 1432 and 2143. Lower right quarter comprises of those weight combinations in which difference between linkability scores for negative data (Type I) and negative data (Type II) ($diff_{ND1-ND2}$) is high, for instance weight combinations 4231, 4321 and 4312. Ideally, we would like weight combinations from upper right quarter, for which we find that weight combinations 3241 and 2341 are the most appropriate, linkability scores for all datasets for these weight combinations are drawn in Fig 4.3 and Fig 4.4, respectively.

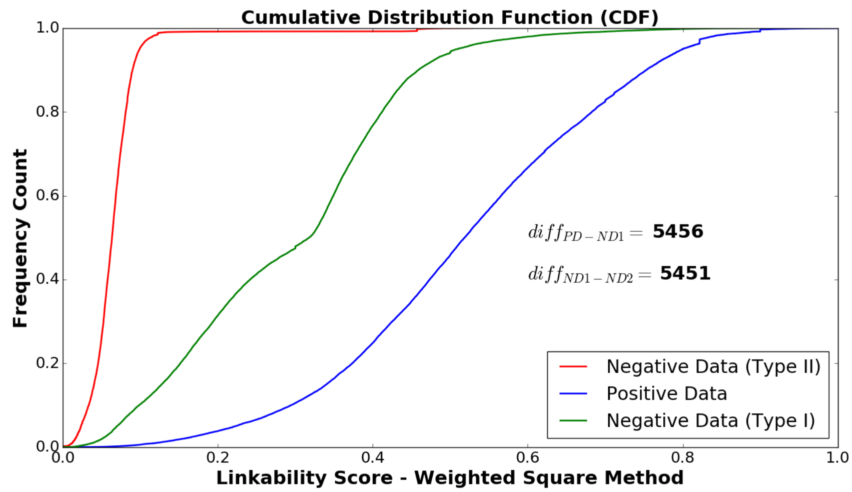



Figure 4.3: Comparison of linkability scores for all types of datasets with features weights as 3, 2, 4 and 1 for four features namely username, name, geo-location and website/url, respectively.

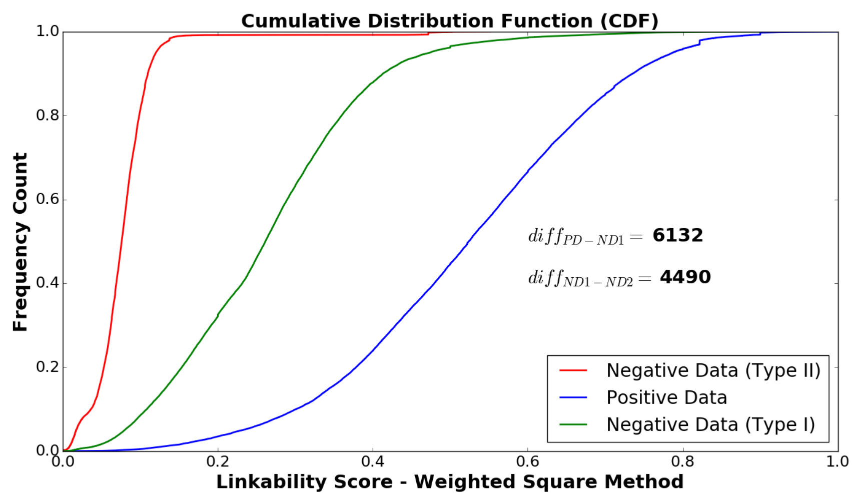



Figure 4.4: Comparison of linkability scores for all types of datasets with features weights as 2, 3, 4 and 1 for four features namely username, name, geo-location and website/url, respectively.

As is evident from both Fig 4.3 and Fig 4.4 that geo-location feature is the most important feature (weight = 4) to distinguish negative data (Type I) identities from positive data identities while website/url is the less important feature (weight = 1), taking all features into consideration. Also observe that Fig 4.4 offers the maximum difference between positive data and negative data (Type I) linkability scores, thereby, ensuring that negative data (Type I) identity would not be mistaken to be as positive identity.

### 4.4.2 Linkability Scores - Probabilistic Method

Results for linkability score computation using probabilistic method are described in this section. We first analyse the effect of individual features on linkability scores. So, in Fig 4.5 and Fig 4.6, we plot the distribution of linkability scores for different types of data (positive and negative Type I & Type II) taking into account only 'username' and 'name of user' as features, respectively.

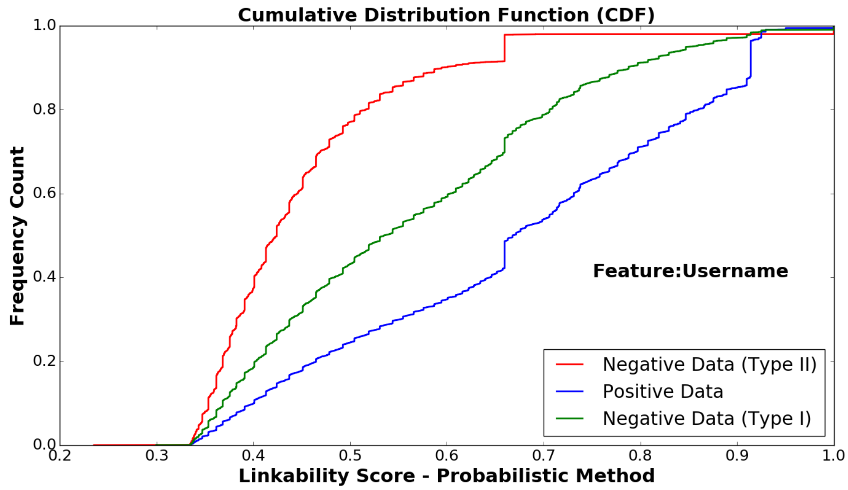



Figure 4.5: Comparison of linkability scores for all types of datasets taking only username as a feature.

It is evident from Fig 4.5 that distribution of linkability scores for both positive data and negative data (Type I) are quite *close* to each other which means that it would be difficult to distinguish between these two types of data. Linkability scores for negative data (Type II) is lesser than the other two types of data which is desirable.

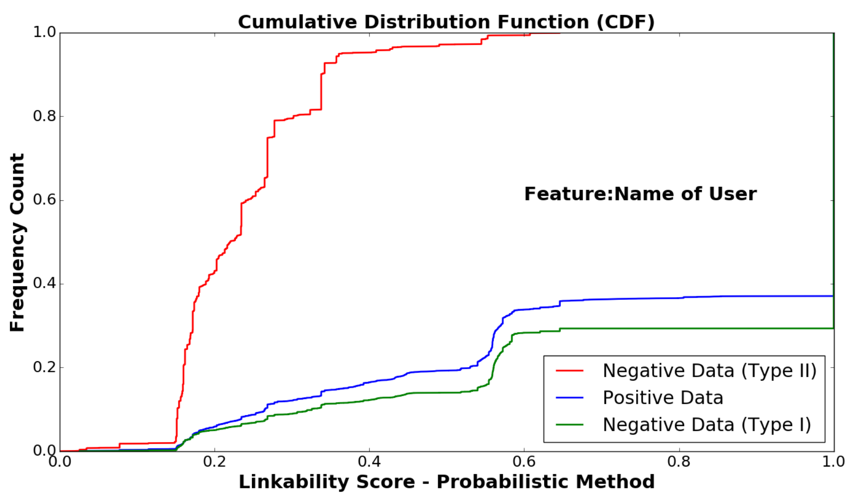



Figure 4.6: Comparison of linkability scores for all types of datasets taking only name of user as a feature.

From Fig 4.6 as well, we can draw similar conclusion, only in this case the distribution between linkability scores for positive data and negative data (Type I) is even more lesser. Negative data (Type II) has its linkability scores at much lower values (mostly less than 0.2) which is desirable. The distribution of linkability scores for all types of data taking all features into consideration is also quite similar to observations made earlier which suggests that probabilistic method would be unable to distinguish between positive data and negative data (Type I). This would possibly be due to the fact that while collecting negative data (Type I), we collected identities which have *similar* name and therefore, these limited number of profile attribute based features are unable to distinguish between positive data and negative data (Type I).

## 4.5 Evaluation Approach

To evaluate the validity of linkability scores computed using weighted sum method and probabilistic method, we reformulate our problem as a **classification problem** in the following steps:-

(1) Threshold Selection: First step is to select an appropriate *linkability score threshold* ($LS_{th}$) by inspecting the distribution of linkability scores.

(2) Classification: Identities whose linkability score is greater that $LS_{th}$ are *classified* to be belonging to same user and those with linkability score less than $LS_{th}$ are *classified* to be belonging to different users.

(3) Evaluation Metrics: Finally in this step, we compute the standard evaluation metrics for classification problem. We present their definition in the context of our problem.

  (a) Precision: It refers number of identities which actually belong to same user from among those identities which have been classified as belonging to same user.

$$Precision = \frac{TP}{TP + FP}$$

  where

$$TP : \forall\ i\ ,\ LS_i\ >\ LS_{th}\ \&\ i\ \in\ D_{PD}$$

$$FP : \forall\ j\ ,\ LS_i\ >\ LS_{th}\ \&\ j\ \notin\ D_{PD}$$

  (b) Recall: It refers number of identities from our entire dataset which have been correctly classified as belonging to same user.

$$Recall = \frac{TP}{TP + FN}$$

  where

$$TP : \forall\ i\ ,\ LS_i\ >\ LS_{th}\ \&\ i\ \in\ D_{PD}$$

$$FN : \forall\ j\ ,\ LS_i\ <\ LS_{th}\ \&\ j\ \in\ D_{PD}$$

  (c) Accuracy: It refers number of identities correctly classified as belonging to either same or different users from among the total identities under study.

$$Accuracy = \frac{TP + TN}{P + N}$$

  where

$$TP : \forall\ i\ ,\ LS_i\ >\ LS_{th}\ \&\ i\ \in\ D_{PD}$$

$$TN : \forall\ j\ ,\ LS_i\ <\ LS_{th}\ \&\ j\ \in\ D_{ND_1+ND_2}$$

$$P : \forall\ k\ ,\ k\ \in\ D_{PD}$$

$$N : \forall\ l\ ,\ l\ \in\ D_{ND_1+ND_2}$$

### 4.5.1 Weighted Sum Method

We present the results obtained for weighted sum method for computing linkability scores. In Fig 4.7, we present accuracy obtained for various threshold values ($LS_{th}$) for linkability scores.

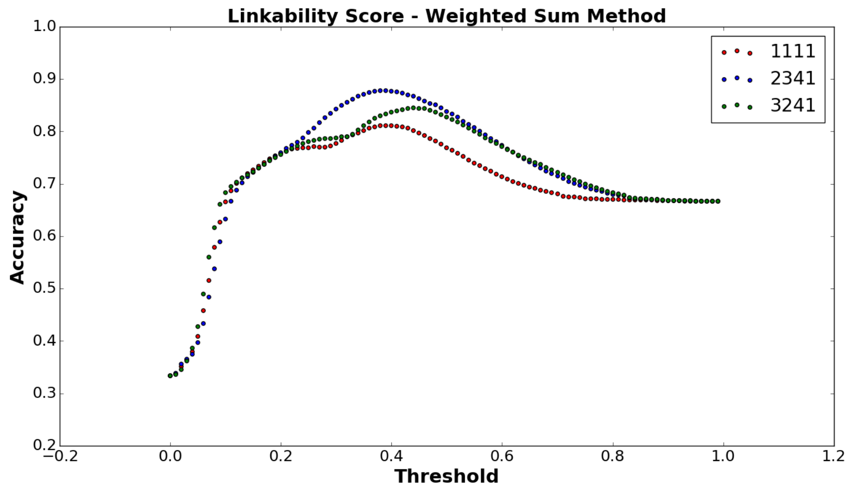



Figure 4.7: Accuracy Vs Threshold, threshold value 0.39 gives the best accuracy of 87%, three feature weight combinations (1111, 2341 and 3241) have been used in the calculations.

Three combinations of feature weights are used for plotting the graph are 1111, 2341 and 3241. Threshold value of 0.39 gives best results in terms of accuracy (87%) when feature weights used are 2, 3, 4 and 1 for username, name of user, location and website features, respectively.

Next, we plot the ROC curve (Fig 4.8) which depicts that best results would be obtained for the feature weight combination of 2, 3, 4 and 1.

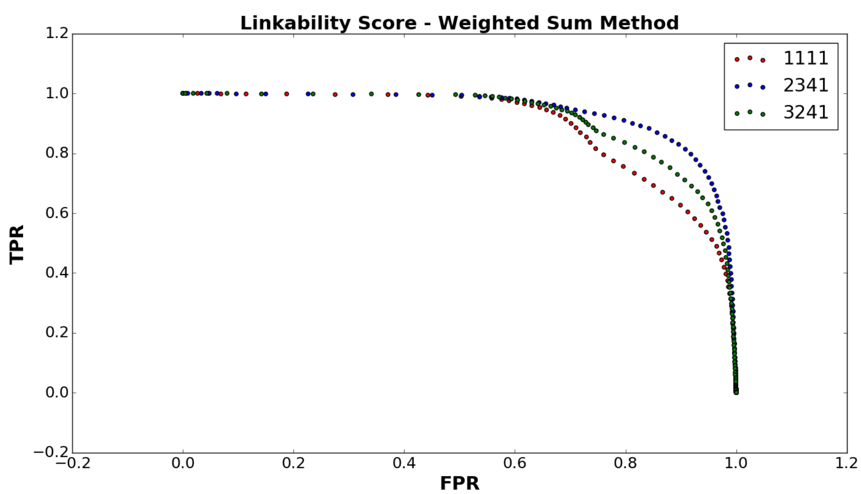



Figure 4.8: False Positive Rate Vs True Positive Rate, i.e. area under the ROC curve shows that weight combinations of 2341 would give best results.

Finally, to compare precision with recall, we plot Fig 4.9 that shows the relationship between precision and recall rate.

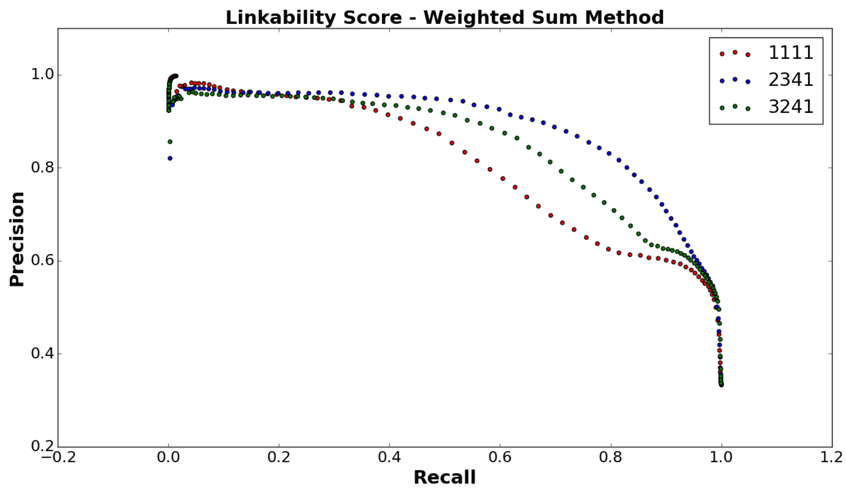



Figure 4.9: Recall Vs Precision.

### 4.5.2 Probabilistic Method

Results for probabilistic method are not as encouraging as those obtained for weighted sum method. This can be due to the fact that probabilistic approach is not able to distinguish between positive data and negative data (Type I) as was observed in Fig 4.5 and Fig 4.6. Plot of accuracy with respect to varying values of linkability score threshold ($LS_{th}$) is drawn in Fig 4.10.

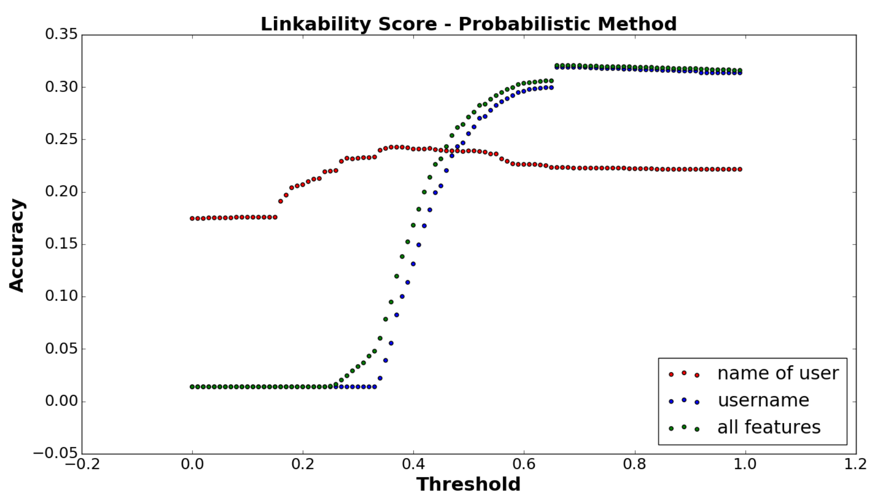



Figure 4.10: Accuracy Vs Threshold, threshold value 0.71 gives the best accuracy of 32%, three scenarios (only 'username' feature, only 'name of user' feature and all features) have been considered.

As observed, the accuracy is merely 32% for a threshold value of 0.71 which suggests that even at a higher threshold value, accuracy is not increased. This suggests the profile attributes alone are not good factors for computing linkability score and that probability approach requires further improvement.

## 4.6 Limitation of Baseline Approach

The probabilistic method did not produce the anticipated results with quiet low accuracy. The weighted sum method produced an accuracy of 87% for a threshold value of 0.39. However, both, weighted sum and probabilistic methods employ a small set of features namely name, username, geo-location and website. Also, these methods fail to capture user's content sharing behavior. So, we move towards state-of-the-art identity resolution methods discussed in the following chapter.

# Chapter 5

# Reformed Linkability Score Computation and Linkability Nudge

We have seen in the last chapter that the proposed methods for linkability score computation, namely Weighted Sum method and Probabilistic method, exploit only a small set of features and fail to capture user's behavior. Also, we infer that Weighted Sum method has higher accuracy as compared to Probabilistic method. Consequently, in this chapter we employ Weighted Sum method over the well known Identity Resolution techniques to compute linkability scores.

## 5.1 Linkability Score

Linkability score quantifies the degree of closeness or separation between two identities on a pair of OSNs. We compute linkability score in three ways based on well known identity resolution methods of NEMO [10], HYDRA [18] and MOBIUS [7] as explained in this chapter.

Our approach to solution is to compute a function that takes a user's identities $i_A$ and $i_B$ on two OSNs $A$ and $B$, respectively as input and compute linkability score between them as below.

$$LS_{i_A,i_B} \leftarrow f_{linkability\_score}(i_A, i_B) \tag{5.1}$$

Identity of a user $u$ on OSN platform $X$ is modeled as feature vector that is values $< v_1^X, v_2^X, ..., v_n^X >$ corresponding to $n$ features $< f_1^X, f_2^X, ..., f_n^X >$. Given an identity pair $< i_A, i_B >$ as input, the function for computing the linkability score is weighted sum of appropriate *feature similarity metric* ($FSM$) between corresponding feature values of identity pair.

$$f_{linkability\_score}(i_A, i_B) \leftarrow \frac{1}{n}\sum_{i=1}^{n} FSM(v_i^A, v_i^B) \tag{5.2}$$

In addition, we also *rank* features based on their contribution to the linkability score. Both linkability score and ranked feature information would be useful to a privacy conscious user who would like to keep linkability score to a lower value. Subsequently a system referred as *linkability nudge* is designed, developed and evaluated which makes soft paternalistic interventions (nudges) whenever a user behavior causes linkability score to go beyond the desired range of linkability score.

### 5.1.1 Architecture

In order to compute linkability scores between pair of identities of a user, we designed a web based application based on Django framework.[4] Fig 5.1 depicts the flowchart of the steps which

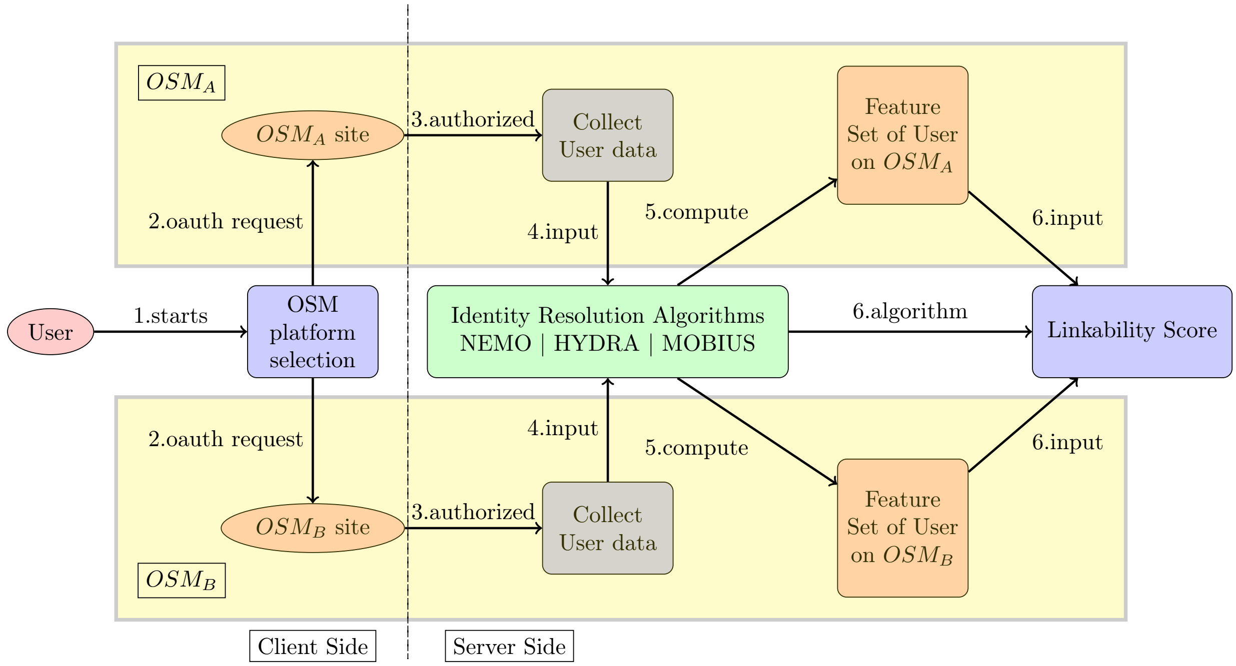



Figure 5.1: Flowchart depicting the steps involved for computing linkability scores.

are performed for computation of linkability scores. On *client side*, there are two key steps as below.

(1) User selects OSN platform, in our experiments as we shall discuss later, the options are Facebook, Twitter and Instagram.

(2) User sends request for grant of access token so that our web application get access to user's data.

On *server side*, following steps are performed.

(3) After obtaining access authorization, web application collects user's data from the OSN platform's API endpoints.

(4) Collected user data is passed as input to identity resolution algorithms which specifies various features, say $< f_1^X, f_2^X, ..., f_n^X >$.

(5) Using the user's data, values of these features are computed on different OSNs, say $< v_1^X, v_2^X, ..., v_n^X >$.

(6) Finally, using the feature vectors and algorithm (namely, NEMO, HYDRA and MOBIUS), linkability scores for each pair of OSNs are computed (refer Fig 5.2) using eq. 5.1 and eq. 5.2.

Given that we are accessing user's data, we have taken utmost care that we follow the principles of ethical research. Through the development of our proposed system of linkability nudge, we are nudging users whenever their actions result in making linkability go beyond desired range. Privacy conscious users are expected to configure their desired range of linkability scores to lower values, by

[4] Django Framework, https://www.djangoproject.com/

way of which nudging will prevent them from inadvertent disclosures, thereby being beneficial. The data which we collect is obtained using temporary access tokens which would typically expire after few hours and we would no longer be able to get user data anytime in future unless user explicitly refreshes them. All users who were involved in evaluation of our nudge were informed about data collection and data usage upfront, they were recruited in evaluation study voluntarily.

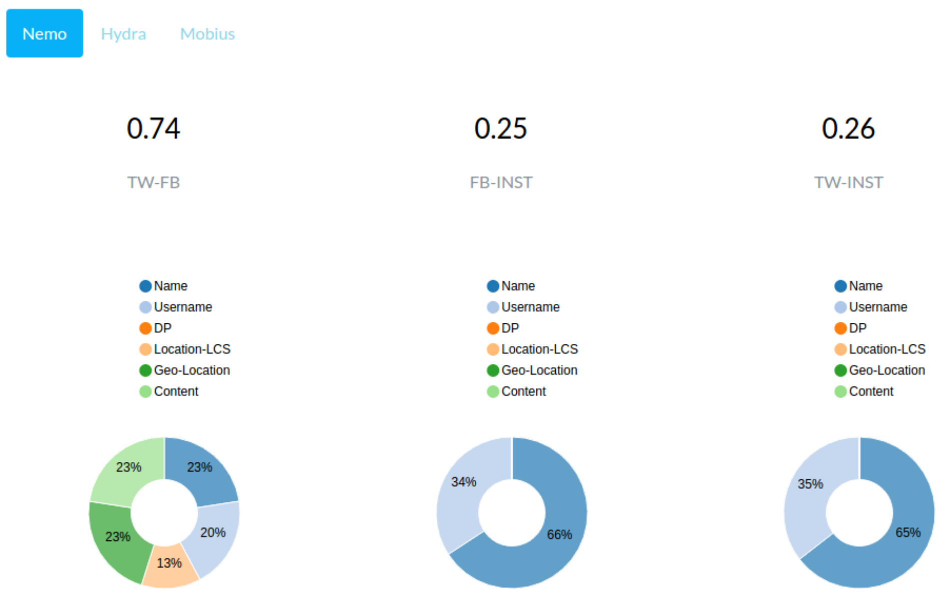



Figure 5.2: Interface where user is shown linkability scores for all pairs of OSNs. Also displayed (pie-chart) is the contribution of factors (attributes) in computation of linkability score.

It is observed from Fig 5.2 that along with linkability scores, we display in pie-chart the contribution of various attributes towards the linkability scores. For instance, in the linkability score of 0.74 between Twitter and Facebook using identity resolution method as NEMO [10], is due to similarities in name, username, location and content posted by user on Twitter and Facebook. Therefore, if the user desires to keep the linkability scores between his identities on Twitter and Facebook to a lower value, then this pie-chart can help him identifying the factors that must be attended to lower linkability score.

### 5.1.2 Methodology

We leverage three well known Identity Resolution (IR) methods namely NEMO [10], HYDRA [18] and MOBIUS [7] to compute linkability score. It may be observed that all these methods propose techniques using user's profile attributes and behavior in order to resolve user's identities across OSNs. However, our aim is to build upon these existing IR methods and propose a metric which we refer as *linkability score*, which quantifies the possibility of linkability or non-linkability of user's identities across OSNs.

Username is a publicly accessible profile attribute of a user which uniquely identifies him within a social network. A user can choose a username, which may be a compressed form of her name or any nickname. Research shows that around 40% users tend to keep same or similar username across social networks, therefore for such users, username can be used to resolve identities across social networks [28, 29]. We verified this in our pre-study and found that nearly 56% respondents said that they have similar usernames across OSNs and 12% respondents wanted to have a similar

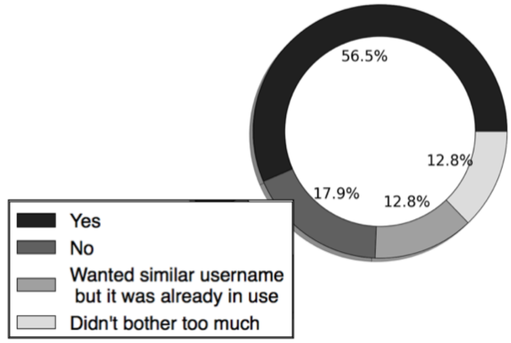


Figure 5.3: Participant responses when asked if they had similar username across two or more social networks.

username but it was already taken as shown in Fig 5.3.
Thus, we seleted a state-of-the-art identity resolution method that concentrates on username and leverages behavioral patterns in username selection referred as **MOBIUS** and proposed by Zafarani et. al. [7]. We have computed the features identified by Zafarani et. al to be most useful for username matching in content of identity resolution, as shown in Table 5.1 and using these features alongwith prior information (labeled data [5] ), we have learned an identification function by employing conventional regression technique drawn from the field of machine learning as shown in Fig 5.4

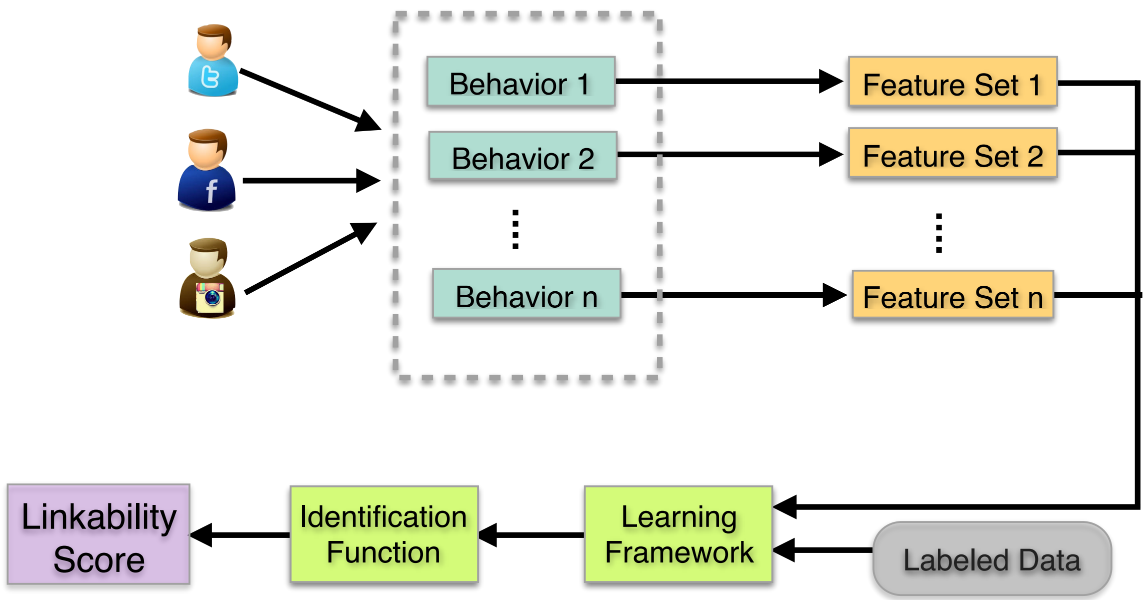


Figure 5.4: Architecture Diagram for calculating Linkability Score using MOBIUS.

In the second IR method used, referred as **NEMO**, Jain et. al. [10] have used four algorithms for identity resolution, namely profile search, content search, self-mention search and network search. In our work, we have used only *profile search* and *content search* algorithms. For computing linkability between two identities on different OSNs, we have considered five features as shown in Fig 5.5 along with their suitable metric for similarity measurement.

[5] Dataset having users' usernames on Facebook, Instagram, Tumblr, Flickr and LinkedIn

Table 5.1: Feature list along with their suitable metrics taken from Mobius as identity resolution method, note that only similarity metric are specified since Mobius is based on only one attribute i.e. 'username'.

| **Behavioral Pattern** | **Similarity Metric** |
|---|---|
| Username Modification | Standard deviation of normalized edit distance between the candidate username and prior usernames. |
| Username Modification | Standard deviation of normalized longest common substring between the username and prior usernames. |
| Username Uniquesness Likelihood | Uniqueness of prior usernames. |
| Using Same Usernames | Exact match. |
| Limited Alphabet | Jaccard similarity between the alphabet distribution of the candidate username & prior usernames. |
| Typing Patterns | Standard deviation of the distance travelled when typing prior usernames using QWERTY keyboard. |
| Username Modification | Standard deviation of the longest common substring between the username and prior usernames. |
| Username Modification | Median of the longest common subsequence between the candidate username and prior usernames. |

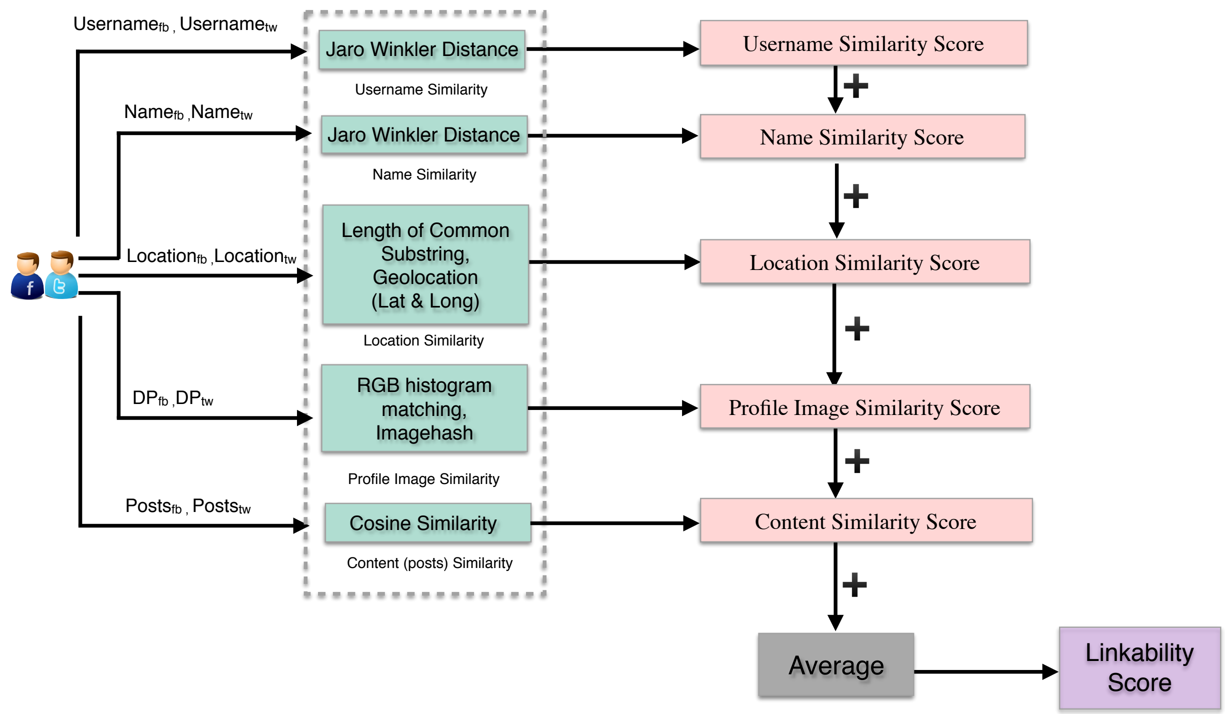



Figure 5.5: Architecture Diagram for calculating Linkability Score using NEMO.

In third IR method that we use, referred as **HYDRA**, Liu et. al. [18] have mainly considered user behavioral modeling, namely, User Attribute Modeling, User Style Modeling and Multimedia Content Generation. User Attribute Modeling considers textual attributes and visual attributes configured in their identities by user on different OSNs, details of those used by us are shown in Fig 5.6.

Finally, it may be stated here that due to restrictions in the endpoints offered by APIs and the number of attributes offered by OSNs namely Facebook, Twitter and Instagram,we could use only

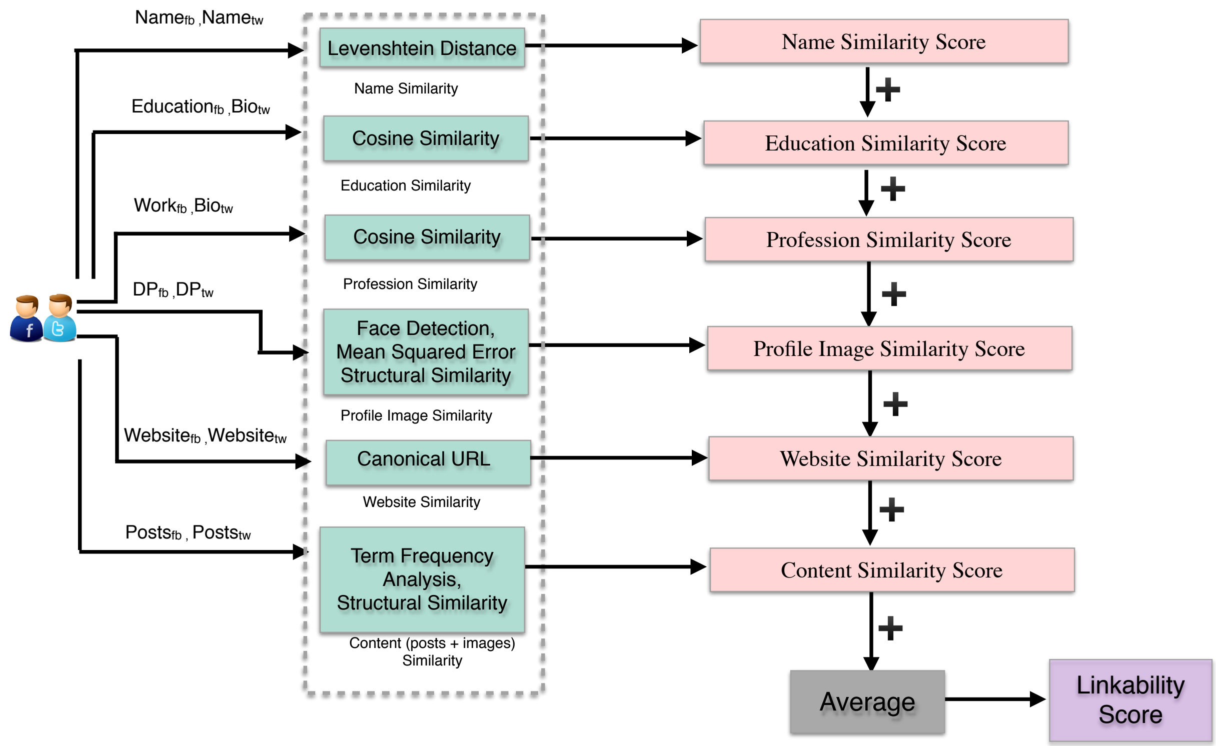



Figure 5.6: Architecture Diagram for calculating Linkability Score using HYDRA.

limited features of NEMO, HYDRA and MOBIUS. Further, different OSNs have varying number of profile attributes (features) for their users, for instance, Facebook has a very large set of user's profile attributes like employment details, school details, etc. but most of them are not present on Twitter or Instagram. Instead, for instance, Twitter has 'about me' where a user **at times** mentions his/her employment and/or school details.

## 5.2 Linkability Nudge

Having explained the computation of linkability score, we describe the design of linkability nudge, a system which introduces soft paternalistic interventions to user whenever user's behavior causes linkability score to change beyond the desired range configured by the user.

### 5.2.1 Design Features

Inspired from the works of Schaub et. al. [26] and Acquisti et. al. [27] for designing privacy notices and nudges, respectively, we adopted the following design features in developing our proposed system of linkability nudge:-

(1) *User Centric Design*: Users are provisioned with the facility to provide their desired linkability score ranges between each pair of OSNs. This ensures that users can *fine-tune* linkability requirements for every pair because there could be a situation that a user wants to keep linkability between OSN $A$ and $B$ low while that between OSN $B$ and $C$ high.

(2) *Opt-out*: We provide an option to users for opting out of the linkability nudge. In this case, users are not exposed to nudge at all and they continue with their OSN usage as if linkability nudge is non-existent. This is important because we have assumed that user

is privacy conscious which may not always be true for that user, so this option is useful for users who are unconcerned about their personal activities getting exposed owing to linkability.

(3) *Balanced Nudging frequency*: In order to be an effective tool, we ensure that users are not too bothered with frequent nudges at every activity that they perform on OSN platform neither do we go to the extreme other side where users are rarely nudged. A balance between the two ends is maintained by ensuring that users are nudged only when any activity that they perform on OSN platform result in change in linkability score beyond their desired pre-configured range.

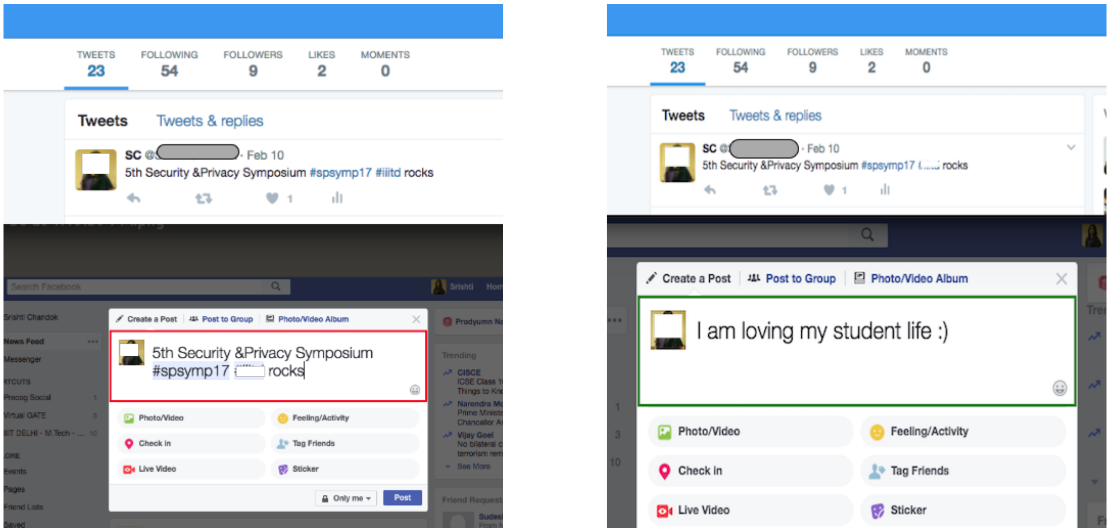
(4) *Succinct Information exposure*: In nudge design managing information exposure to the end user is important parameter to be considered so that user is not overwhelmed with lot of information. Keeping this in view, we only show short messages in the nudge notification and only the score values in nudge pop-up.

### 5.2.2 Nudge Design

There are numerous ways in which nudges can be designed. In our proposed nudge design, we have focused on two types of nudges.

(1) **Content-driven Color Nudge**: Users having identities across multiple OSNs often indulge in *cross posting* which means posting same or similar information across multiple OSNs. Such behavior increases similarity in their identities, thereby increasing the linkability score. Our first nudge design addresses this particular issue by nudging the user through use of color. Whenever user types a post which is similar to any of the existing posts made by the user on other OSNs, we nudge the user by coloring the post's text box border with red as show in Fig 5.7a. Color is green as long as linkability scores are within their pre-configured ranges as shown in Fig 5.7b. This is an indication to user that this post is an instance of *cross posting* which is going to increase user's linkability across OSNs. Nudge being only a soft paternalistic intervention, we leave the text box colored with red and let user decide whether user wants to continue making the post or refrain from making the post.

(2) **Attribute-driven Notification Nudge**: User with multiple identities across OSNs maintain their identities with that there is overlap among the values of attributes specified by them on these OSNs. More is the overlap, more similar the identities would be and higher would be the linkability scores. In fact, the initial linkability scores being computed when user grants authorization is mostly due to similarity in profile attributes like name, username, location, profile picture and so on. Whenever user modifies value of any profile attribute over an OSN which causes change in linkability score such that the score goes beyond the pre-configured desired range, then the user is nudged. Nudge is delivered in the form of a pop-up notification on top right of screen with a short message saying *'Your linkability with Facebook has increased'* as show in Fig 5.8. Again here, being only a soft paternalistic intervention, we allow user's change in attribute to take place and let user decide whether user wants to revert the change or not.




(a) Facebook post is similar to Twitter post, the text box around post shows up in red.

(b) Facebook post is different from Twitter post, the text box around post shows up in green.

Figure 5.7: Illustration of Content-driven Color Nudge in which it is assumed that user has already made a post on Twitter and then is making a post on Facebook.

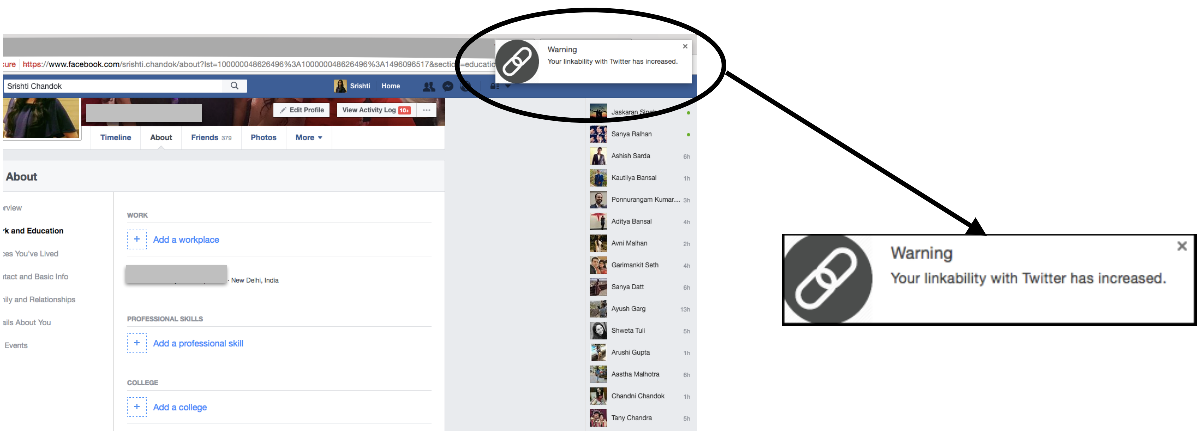



Figure 5.8: Illustration of Attribute-driven Notification Nudge on top right of the Facebook page alerting user with a short message that 'Your linkability with Twitter has increased', similar notifications are present to user on interfaces of Twitter and Instagram. Also shown is the enlarged view of nudge notification.

The browser plugin can be downloaded from the link:
https://drive.google.com/file/d/0B8wqXuB3kl0DS2Y2XzFlbUNjWVk/view?usp=sharing

# Chapter 6

# Nudge Architecture, User Evaluation and Results

## 6.1 Nudge Architecture

We implement linkability nudge by developing a chrome browser plugin that can be installed on user's web browser.[6] This plugin monitors user's behavior in terms of the content being posted over OSNs and changes to profile attributes being made on OSNs. Architecturally, linkability nudge comprises of three main components namely browser extension, nudge server and linkability compute server. Fig 6.1 shows how these components get invoked on different user activities. The three components of linkability nudge are described below.

### 6.1.1 Browser Extension

This is the only component where a user is required to install on Google chrome web browser. It performs a number of functions:

(1) Maintains user's identity and user context across the entire user session.

(2) Captures user's posting activity and changes in profile attributes on all configured OSNs.

(3) Also displays linkability nudge in various forms, discussed later.

### 6.1.2 Nudge Server

This is the component that is required to be installed on server side. It is an intermediary which sits between the *browser extension* and *linkability compute server*. It performs following functions:

(1) Receives user's access token from browser extension and sends them to OSN servers to obtain user's data.

(2) Stores user's data in a database temporarily.

(3) Passes the information pertaining to user's activities like making a post or changing profile attribute to the linkability compute server.

(4) Sends across the newly computed linkability scores to the browser extension from time to time based upon user's activities.

[6] Plugin shall be soon made available on Chrome Web Store for people to use and provide their feedback

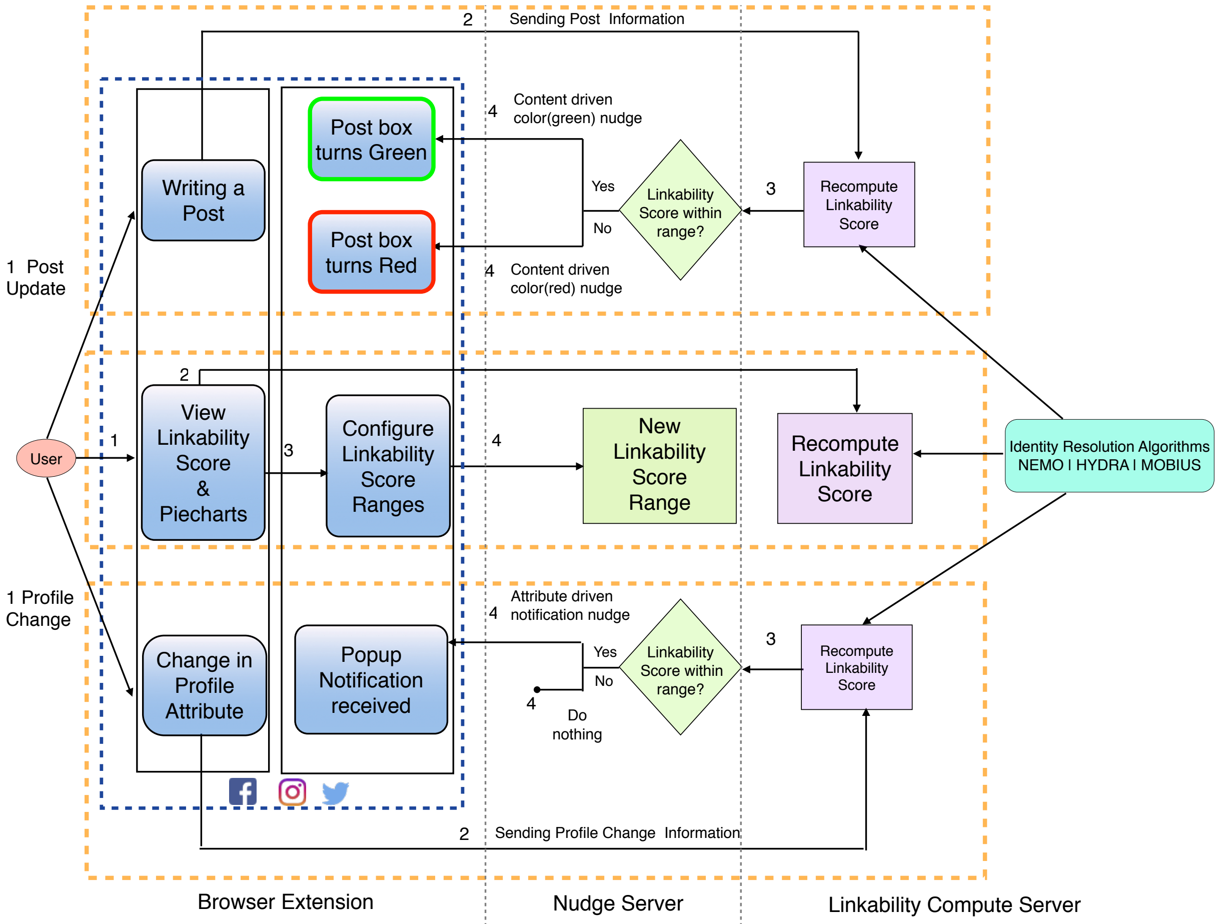



Figure 6.1: Architecture Diagram showing the interaction of user with the system. Boxes show different activities which user performs, namely post update, profile attributes change and viewing linkability scores.

### 6.1.3 Linkability Compute Server

This is the component which performs most of the heavy computation involved in calculation of linkability scores and it is to be installed on server side. It performs following functions:

(1) It implements the identity resolution methods to compute linkability scores.

(2) It retrieves user's data from the database as input to compute linkability scores at initial setup time.

(3) Subsequently, it receives every user activity's information (whether making a post or changing profile attribute), recomputes linkability scores and sends them back to nudge server.

## 6.2 User Evaluation

In this section, we present our approach for evaluating the system of linkability nudge. We follow the standard process of performing controlled lab study.

### 6.2.1 Participants

In order to gauge user's perceptions and opinions with respect to usage and linkability issues in a multi-OSN scenario, we engaged 40 participants in *pre-study questionnaire*. Subsequently, we filtered out and recruited only 12 participants for controlled lab study who had their accounts on all the three OSNs (namely Facebook, Twitter and Instagram) on which our proposed linkability nudge was designed. Participants were within the age group of 18-26 years, with 67% female and 33% male comprising of mostly undergraduate students studying computer science. Demographics of participants are provided in Table 6.1.

Table 6.1: Demographics of participants

| Attribute | Range | Pre-Study | Experiment |
|---|---|---|---|
| **# Participants** | | 40 | 12 |
| **Age** | 18 - 20 | 4 | 3 |
| | 21 - 23 | 11 | 5 |
| | 24 - 26 | 25 | 3 |
| **Gender** | Male | 18 | 4 |
| | Female | 22 | 8 |
| **Highest Degree** | 12th Class School | 7 | 4 |
| | Graduate | 13 | 5 |
| | Postgraduate | 19 | 3 |
| **Discipline** | Computer Science | 29 | 11 |
| | Science | 4 | 0 |
| | Commerce | 3 | 0 |
| | Others | 4 | 1 |
| **Occupation** | Student | 21 | 12 |
| | Private Job | 13 | 0 |
| | Service | 2 | 0 |
| | Self Employed | 4 | 0 |

### 6.2.2 Study Design

We conducted controlled lab study in two phases namely

- *Control Period*: Participants are not exposed to linkability nudge. They are asked to perform tasks as outlined in next section for around half an hour.
- *Treatment Period*: Participants are subjected to linkability nudge. In this phase again, we ask the participants to perform the same tasks as performed in control period for around half an hour.

### 6.2.3 Tasks

In order to prompt user to perform some activities so that role of linkability comes into play, we designed two types of tasks: (a) Making *scenario based posts* and (b) Changing *profile attributes* for the identities maintained by users on OSNs. Following instructions regarding the task details were provided to participants of user study.

(1) **Scenario based posts** Please make at least around 8-10 posts, you can delete the posts after the study. For each of the post,

- You are free to select media of your choice
- You are free to post over one or more platforms

(a) Official Post: Suppose that you attended a technical presentation (or some recent events), make a post about the same on social media platform(s).
- You are free to post as much details as you would like (say location, speaker, organization, etc)

(b) Personal Post: Suppose that you planned a vacation with your family or by yourself alone, make a post about the same on social media platform(s).
- You are free to post as much details as you would like (say people with whom you plan your vacation, destination, mode of transport, etc)

(c) Hangout: Suppose that you make an impromptu hangout with your friends, make a post about the same on social media platform(s).
- You are free to post as much details as you would like (say people with whom you plan your hangout, destination, mode of transport, etc)

(d) Sharing: Users in your network on a social media platform are making posts, select **any** platform, **any** user in your network and share **any** of his/her post.
- You are free to **select** visibility of the shared post
- You are free to include your opinion while sharing the post

(e) Self Promotion: Suppose that you have achieved something or there is an important update in your life which you want to share and promote yourself.

(f) Location Sharing: Make a post in which you reveal your geo-location.

(g) Self Disclosure: Share a post with URL of your own homepage in it or URL of your profiles in other social media platforms.

(2) **Revisiting profile configuration** Explore your basic profile setting pages on all three OSNs which you have and review the values of field. In the process of reviewing, you may leave the field values unchanged or fill the unfilled fields or modify values of filled fields.

More specifically, you may modify / review following attributes:-

(a) Facebook: Location, Website, Education, Work & Display Picture.

(b) Twitter: Name, Location, Website, Bio & Display Picture.

(c) Instagram: Name, Username, Website, Bio & Display Picture.

Participants were asked to perform same set of tasks in both *control period* and *treatment period*, however the key difference is that before the start of *treatment period*, participants were explained about the broader concept of linkability of their identities across OSNs and implications of the same. In addition, participants were also demonstrated the usage of nudge through an instructive video. Particularly, they were instructed that operation of two types of nudges namely *content-driven color nudge* and *attribute-driven notification nudge*, as below.

- **Content-driven Color Nudge**: Whenever you compose a message (or tweet) on these OSNs, if the text of your message changes the linkability beyond the desired linkability score, then the border of text box will become red in color else it will be green. While making post, please be patient, because every character pressed is being sent back to server and scores are being recomputed on the fly.

- **Attribute-driven Notification Nudge**: Whenever you make change in any attribute listed in tasks which results in change in the linkability beyond the desired linkability score, then a pop-up notification alert on top-right of the page will appear.

In addition to the earlier tasks, during *treatment period*, participants were informed that they may visit the linkability score configuration page and re-adjust the various ranges of linkability score between different pair of social networks, keeping the following rule of thumb.

- Decrease the range between a social network pair if you want your identities on that pair to be as separate as possible (low linkable).

- Increase the range between a social network pair if you want your identities on that pair to be as similar as possible (highly linkable).

Additionally, participants were informed that they may visit the home page of linkability nudge as many times as they want to view the new values of linkability scores between all pairs of social network. Also, they may see the pie charts for all pairs of OSNs that they have connected on the configuration page. These pie charts show the contribution of each profile attribute or content towards their linkability score. Participants were explained that these pie-charts could help you to decide as to which attribute to alter in order to approach towards their desired linkability score.

## 6.3 Results

Here we compare user behavior with respect to the tasks being performed in the two phases namely *control period* and *treatment period*.

### 6.3.1 Implications of Nudge

In order to understand the extent to which our proposed system of linkability nudge created awareness among participants, we asked few follow-ups questions in *post-study questionnaire*. 58% of the participants understood the broad concept of linkability score either completely or most of it, while 42% said that they understood a little bit about it as shown in Fig 6.2. 42% of participants

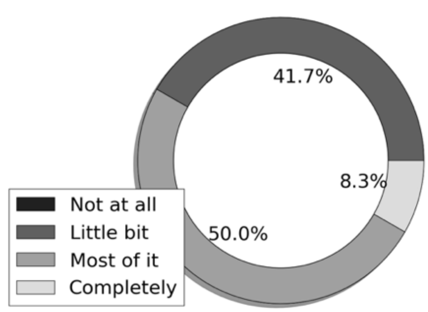



Figure 6.2: Participant responses when asked about understanding of the broad concept of linkability score.

said that they are absolutely sure that they are more aware and informed about the linkability of your multiple identities across OSNs after using the nudge while another 42% responded by saying 'somewhat' more informed, as shown in Fig 6.3. Most of the participants (84%) said that they did

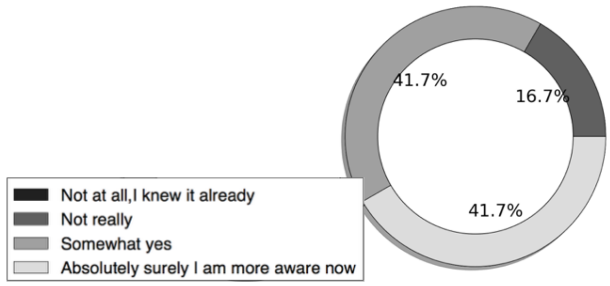



Figure 6.3: Participant responses when asked about awareness of the linkability of their multiple identities across OSNs.

notice the factors contributing to their linkability scores which itself suggest that participants were well informed about the causes for their linkability scores, as shown in Fig 6.4.

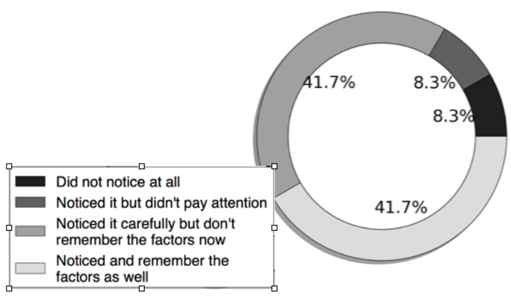



Figure 6.4: Participant responses when asked about noticing the factors contributing to linkability scores on different pairs of OSNs.

### 6.3.2 Interactions with Nudge

The time line plot helped us in understanding user's interactions with linkability nudge (degree of participation) and vice-versa (nudging frequency).

*Degree of Participation*: Based on the amount of time spent and number of tasks performed (shown in Fig 6.6) both during control and treatment period, we can divide participants among three categories. P1, P3 and P6 performed at least 8 or more tasks, taking into account both scenario based posts (shown in + symbol) and profile changes (shown in $\times$ symbol) during treatment period, we consider them *highly active*. While P4 and P5 also spent entire duration of one hour but they performed very less number of tasks during treatment period. P10 and P12 performed reconfigurations in their linkability scores (shown in $\star$ symbol) and were *moderately active*. While the remaining participants performed at least two tasks and were *least active*. We also recorded passive activities of participants in which they viewed their linkability scores (shown in $\triangleright$ symbol) and factors contributing to those scores in form of piechart (shown in $\circ$ symbol).

*Nudging Frequency*: Participants were nudged during treatment period while during control period, they were not nudged (in Fig 6.6, transition from control to treatment period is depicted by a |

symbol). Content-driven color nudge is depicted by either $\bigtriangledown$ (red) symbol or $\triangle$ (green) symbol while Attribute-driven notification nudge is depicted by $\square$ symbol. Participants who were *highly active* were also nudged the most, more specifically P6, P3 and P1 received nudges 7, 13 and 10 times, respectively. Participants who were *moderately active* received at least twice while the *least active* ones were nudged at least once. Fig 6.5 shows the activities performed by P3. He started with viewing the linkability score distribution through the pie-charts. He then configured the linkability score range. Subsequently, he performed the tasks and made various post and profile attribute changes. During the course of activities in the control period, P3 did not receive any nudge from the extension. After the transition from control to treatment, P3 again configured the linkability score range followed by post and profile changes. P3 made around 10 posts and he observed the red color box for only 1 post. 3 out of 7 profile changes prompted a notification nudge indicating the crossing of linkability score range.

### 6.3.3 Impact of Nudge on User Behavior

We may recall that nudge is an intervention which makes users more informed so that they may take better decisions. By design, nudges are suggestive and not binding on a user. Consequently, we observed that at times users did change their behavior while at other times they overlooked the nudge.

*Impact of Content-driven Color Nudge*: From Fig 6.6, we see that both participants P11 and P12 in their last activities tried to make a post after which they were prompted with a content driven red color nudge (+ symbol followed by $\bigtriangledown$ (red) symbol) and they refrained from making the post. In contrast, participant P1 continued to make a post even when content driven red color nudge was displayed (+ symbol followed by $\bigtriangledown$ (red) symbol which is again followed by + symbol indicating that participant continued to make the post).
*Impact of Attribute-driven Notification Nudge*: From Fig 6.6, we see that participants P6 and P10 performed a profile change which triggered notification nudge which is immediately followed up by them to make change in linkability score range ($\times$ symbol followed by $\square$ symbol followed by $\star$ symbol). In contrast, participant P3 made a number of profile changes and was shown notification nudge which was ignored (in other words linkability score were not reconfigured neither was profile change undone). P12 after having shown notification nudge only viewed linkability scores.

### 6.3.4 Nudge Utility

With respect to utility of nudge, we in *post-study questionnaire* asked participants two specific questions, as below along with the responses obtained.

- *Which aspect/feature of nudge was liked the most by participants?* We recall that our proposed system of calculation of linkability score and linkability nudge had four key features namely a) Notification that would appear on top-right of screen, b) Red color border while posting if user is doing cross-posting, c) Linkability score values for each pair of OSNs and d) Pie-charts depicting contribution of attributes towards linkability scores. As shown in Fig 6.7, both color nudge and pie-charts showing the contribution of profile attributes towards linkability score were equally liked by nearly 83% of the participants.

- *What is overall view of nudge to participants?* To gauge the overall assessment of participants in terms of *usability* of the proposed linkability nudge, we asked them to report

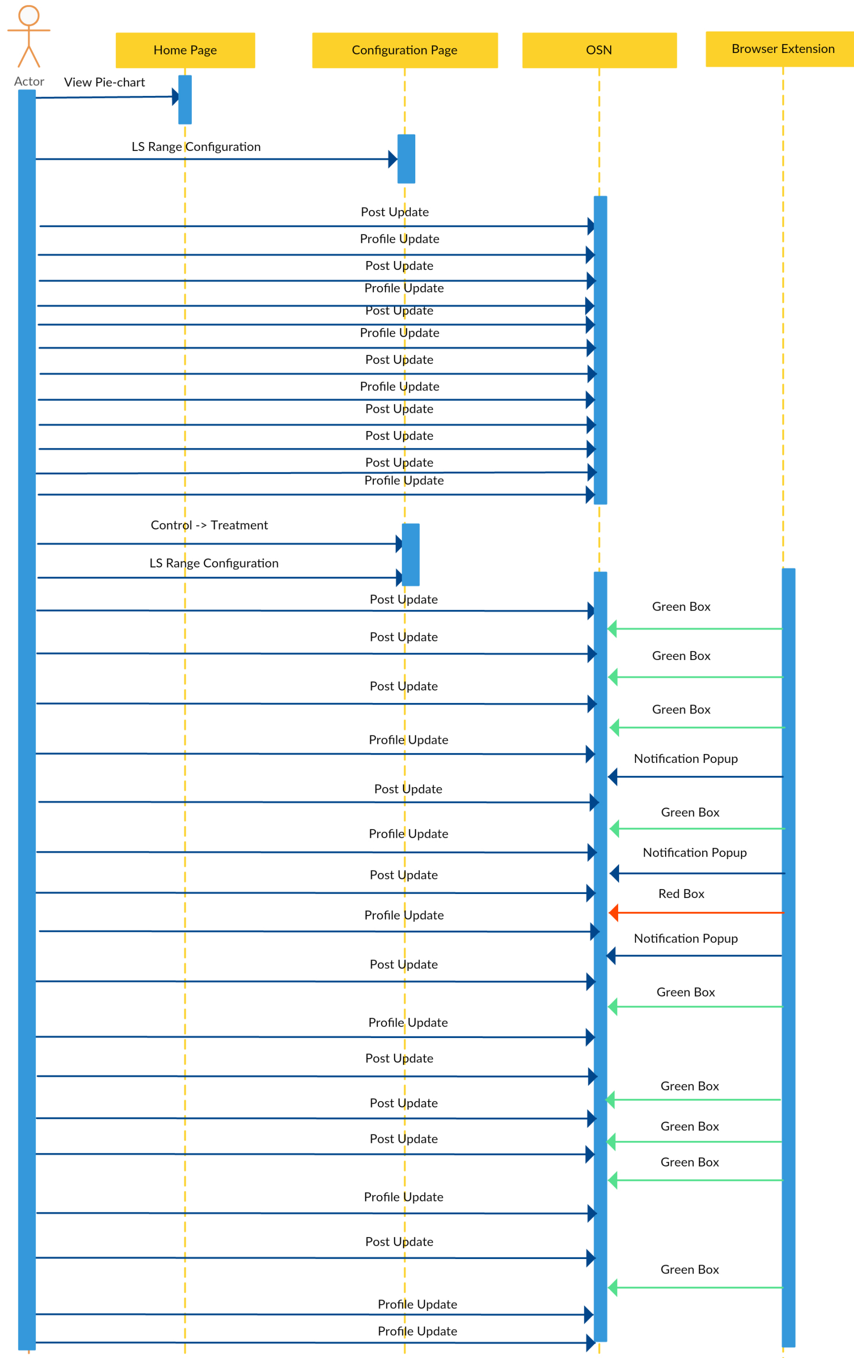



Figure 6.5: Activities performed by participant 3.

whether they found nudge to be useful or not, and if they found it to be useful, how easy it was to use. Fig 6.8 depicts responses of participants in which is evidently clear that most of them found it to be useful, however, some of them also found its usage to be complicated. Some participants expressed concern over complicated usability, on further investigation, we found that it was mainly due to the time delay (2-5 seconds) which they experienced

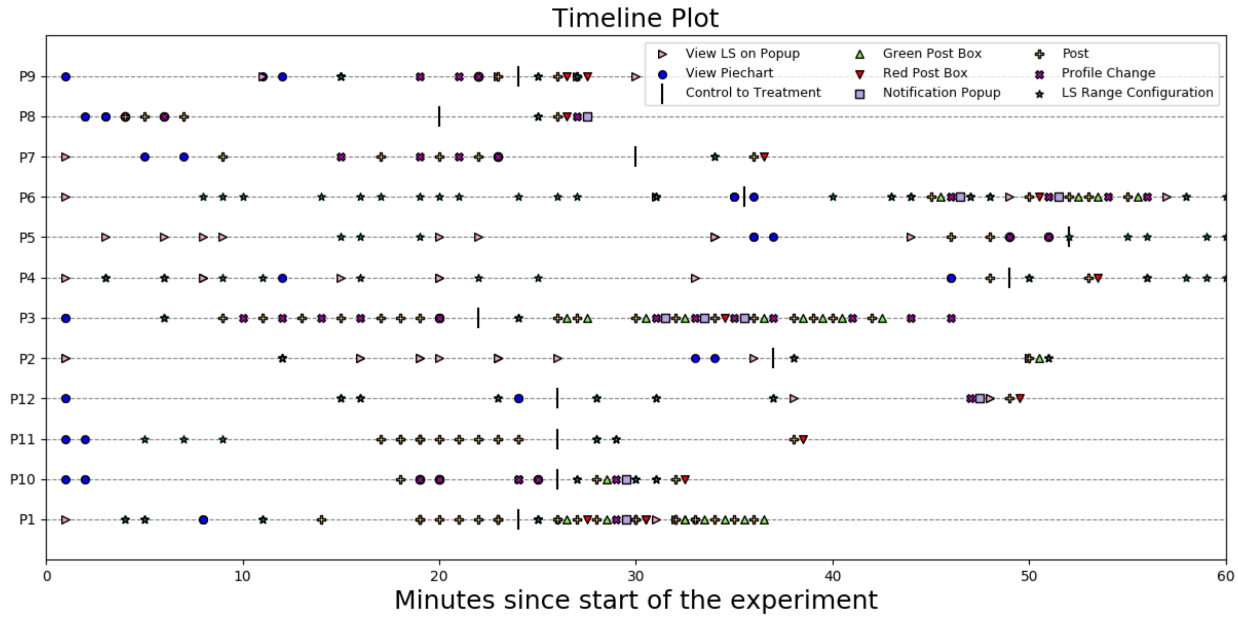



Figure 6.6: Complete timeline of activities of all 12 participants who took part in controlled lab study performing various tasks in control and treatment period.

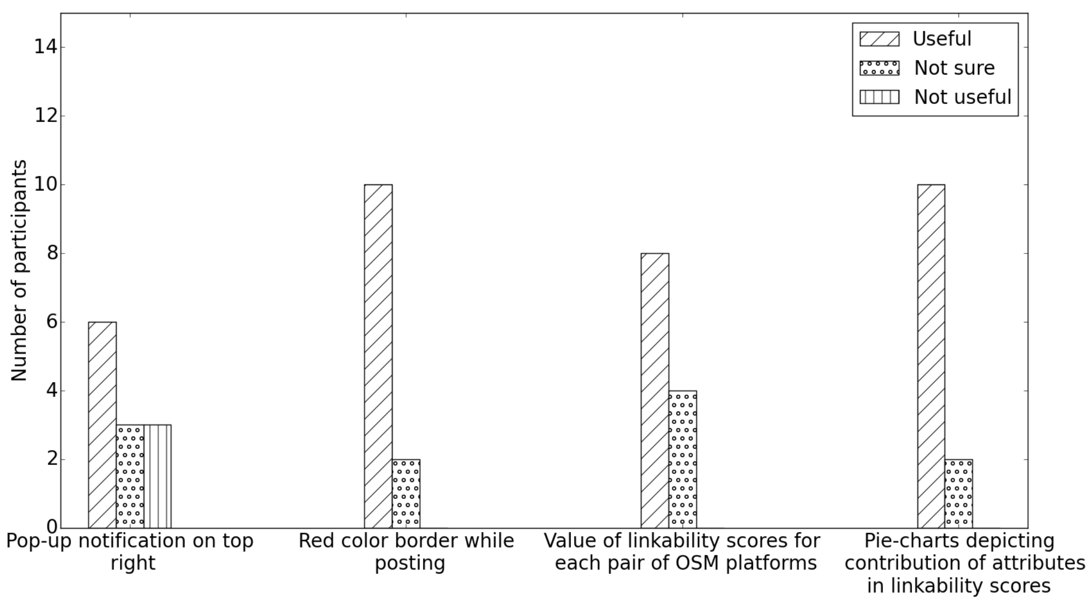



Figure 6.7: Participant responses when asked about usefulness of various features of nudge.

while making post during treatment period. This is because each word typed is sent back to server for re-computation of the linkability score causing the delay.

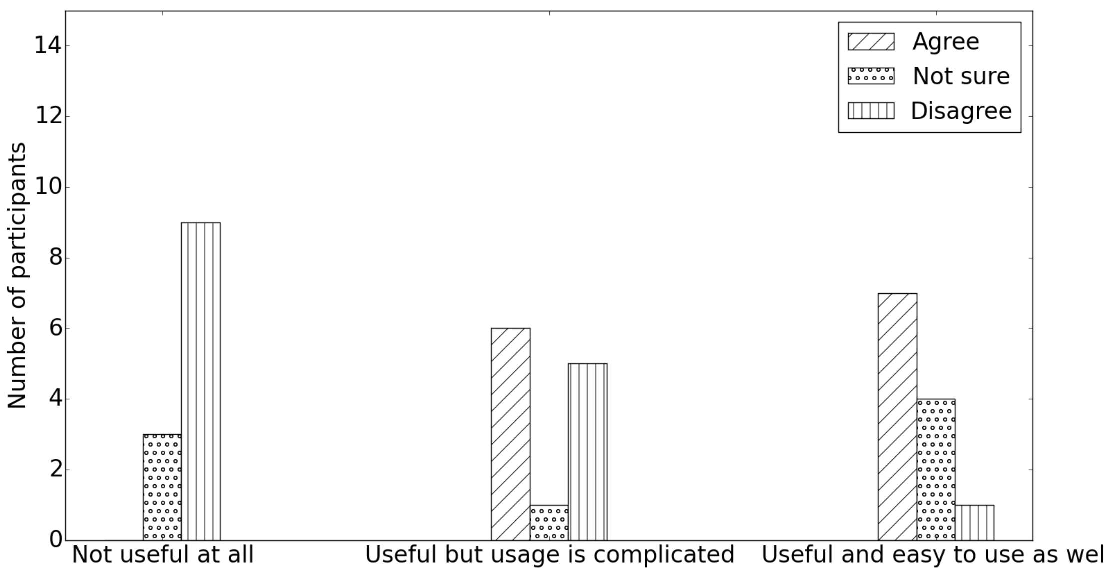



Figure 6.8: Participant responses when asked about the overall view of the linkability nudge.

# Chapter 7

# Discussion, Limitations and Future Work

## 7.1 Discussion

We have developed a system that intends to help privacy conscious users to maintain anonymity across OSNs. The system computes linkability score for every pair of social network configured by the user employing the features from three state-of-the-art methods for identity resolution. The system aims to enlighten the users about contribution of various attributes towards the linkability of accounts. This would help them to decide on the attributes whose values should be changed in order to get the desired linkability score. When a user inadvertently changes any attribute to make it similar to his identity on any other OSN, the system will nudge the user with a short message to indicate the increase in linkability of his accounts. Being a soft paternalistic intervention, the nudge occurs in the form of a pop-up notification, which does not compel the user to alter his decision, instead just warns the user. In this thesis, we proposed and evaluated two approaches for computing linkability scores, namely Weighted Sum method and Probabilistic method. We inferred that Weighted Sum method works better than the Probabilistic method and further used the former to compute linkability scores based on state-of-the-art identity resolution methods. We focus on how to assist privacy conscious users to find and analyse the linkability of their various social networks. For this, we developed a system where a user can connect his Facebook, Twitter and Instagram accounts and the system will return his current pairwise linkability scores alongwith piecharts showing the contribution of various factors leading to that score. Users can configure the desired range of linkability and can choose to opt-out of the nudge.

By conducting the lab study, we came to the conclusion that the users maintaining multiple identities across OSNs were able to see, in quantifying terms, the linkability of their identities between each pair of OSN platform. Linkability nudge helped some of the participants to take corrective measures to avoid inadvertent disclosure of their personal information owing to increased linkability. User evaluation validates that linkability nudge is indeed quite helpful in making users understand the concept of linkability and helps them through soft interventions to remain within their desired linkability ranges. The purpose of linkability nudge was to help users understand the nuances involved in linkability of their identities across OSNs. The goal is that when they perform an activity (making a post or changing profile attribute), they are conscious of the fact that it may increase or decrease linkability of their identity with respect to their identities on other OSNs. Participants of user study exhibited varied level of participation and were intervened by all types of nudge designs during the controlled lab study. Most of the participants liked the color nudge reinforcing the notion that simple designs make significant impact.

## 7.2 Limitations

One of the Identity Resolution methods used, namely MOBIUS, works on usernames. However, Facebook's API has no endpoint for fetching username of the account. So, we have employed scrapping of user's Facebook page using a dummy account to get the username. This method is not ethical and Facebook tends to block the dummy account after sometime. We have tested our system with small number of participants. Also, we expected more active participants and in future we would explore ways to improve it. Behavior of some users changed when they were intervened by nudges while that of others didn't change. We would investigate further to find out the reasons which prompted users to behavior in one way or the other. Linkability nudge was able to make most of the participants more aware of the linkability issues. Some participants expressed concern over complicated usability, on further investigation, we found that it was mainly due to the time delay (2-5 seconds) which they experienced while making post during treatment period. This is because each word typed is sent back to server for re-computation of the linkability score causing the delay.

## 7.3 Future Work

As a part of future work, we can improve on the time taken by the system to collect the data in real-time. In this thesis we considered that all attributes have equal contribution towards the linkability score. We can improve the efficiency of the system by giving different weights to the attributes through learning by the dataset. We shall work to improve the engineering design so as to reduce the delay in re-computation of linkability scores on every activity of the user. Currently our system works for only three social networks, namely, Facebook, Twitter and Instagram. We can add more social networks to our system to make users aware of their linkability across all of them. In future, we plan to deploy our proposed system of linkability nudge in public domain and conduct a field study to understand its impact more extensively.

# Bibliography


[1] Tweet by Cisco Employee. http://www.nbcnews.com/id/29901380/ns/technology_and_science-tech_and_gadgets/t/getting-skinny-twitters-cisco-fatty/#.WUOQqVKB1-U

[2] Facebook post by Georgia high school teacher. https://www.californiabusinesslitigation.com/2013/05/high_school_teacher_files_an_a.html

[3] Identity Resolution Example http://edition.cnn.com/2012/10/31/tech/social-media/sandy-twitter-hoax/

[4] HootSuite web application, https://blog.hootsuite.com/types-of-social-media/

[5] Pew Research Center, Social Media Report, http://www.pewinternet.org/2016/11/11/social-media-update-2016/

[6] Kumaraguru, Ponnurangam, and Lorrie Faith Cranor. *Privacy indexes: a survey of Westin's studies.* Technical report, Carnegie Mellon University CMU-ISRI-5-138, 2005.

[7] Zafarani, Reza, and Huan Liu. *Connecting users across social media sites: a behavioral-modeling approach.* Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2013.

[8] Bartunov, Sergey, et al. *Joint link-attribute user identity resolution in online social networks.* Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining, Workshop on Social Network Mining and Analysis. ACM. 2012.

[9] Carmagnola, Francesca, Francesco Osborne, and Ilaria Torre. *User data distributed on the social web: how to identify users on different social systems and collecting data about them.* Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems. ACM, 2010.

[10] Jain, Paridhi, Ponnurangam Kumaraguru, and Anupam Joshi. *@ i seek'fb.me': identifying users across multiple online social networks.* Proceedings of the 22nd international conference on World Wide Web. ACM, 2013.

[11] Jain, Paridhi. *Automated Methods for Identity Resolution across Heterogeneous Social Platforms.* Proceedings of the 26th ACM Conference on Hypertext & Social Media. ACM, 2015.

[12] Liu, Jing, et al. *What's in a name?: an unsupervised approach to link users across communities.* Proceedings of the sixth ACM international conference on Web search and data mining. ACM, 2013.

[13] Goga, Oana, et al. *Exploiting innocuous activity for correlating users across sites.* Proceedings of the 22nd international conference on World Wide Web. ACM, 2013.

[14] Lim, Bang Hui, et al. *# mytweet via Instagram: Exploring user behaviour across multiple social networks.* 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, 2015.

[15] Zhang, Haochen, et al. *Online social network profile linkage.* Asia Information Retrieval Symposium. Springer International Publishing, 2014.

[16] Li, Jiexun, G. Alan Wang, and Hsinchun Chen. *Identity matching using personal and social identity features.* Information Systems Frontiers 13.1 (2011): 101-113.

[17] Leenes, Ronald. *Context is everything sociality and privacy in online social network sites.* IFIP PrimeLife International Summer School on Privacy and Identity Management for Life. Springer Berlin Heidelberg, 2009.

[18] Liu, Siyuan, et al. *Hydra: Large-scale social identity linkage via heterogeneous behavior modeling.* Proceedings of the 2014 ACM SIGMOD international conference on Management of data. ACM, 2014.

[19] Wang, Yang, et al. *A field trial of privacy nudges for facebook.* Proceedings of the SIGCHI conference on human factors in computing systems. ACM, 2014.

[20] Wang, Yang, et al. *Privacy nudges for social media: an exploratory Facebook study.* Proceedings of the 22nd International Conference on World Wide Web. ACM, 2013.

[21] Wang, Yang, et al. *From facebook regrets to facebook privacy nudges.* Ohio State Law Journal 74 (2013): 1307.

[22] Wang, Yang, et al. I regretted the minute I pressed share: A qualitative study of regrets on Facebook. Proceedings of the Seventh Symposium on Usable Privacy and Security. ACM, 2011.

[23] Ziegeldorf, Jan Henrik, et al. *Comparison-Based Privacy: Nudging Privacy in Social Media (Position Paper).* International Workshop on Data Privacy Management. Springer International Publishing, 2015.

[24] Almuhimedi, Hazim, et al. *Your location has been shared 5,398 times!: A field study on mobile app privacy nudging.* Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, 2015.

[25] Zhang, Bo, and Heng Xu. *Privacy Nudges for Mobile Applications: Effects on the Creepiness Emotion and Privacy Attitudes.* Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. ACM, 2016.

[26] Schaub, Florian, et al. *A design space for effective privacy notices.* Eleventh Symposium On Usable Privacy and Security (SOUPS 2015). USENIX Association, 2015.

[27] Acquisti, Alessandro, et al. *Nudges for Privacy and Security: Understanding and Assisting Users Choices Online.* (October 25, 2016). Available at SSRN `https://ssrn.com/abstract=2859227`

[28] Terence Chen, Mohamed Ali Kaafar, Arik Friedman, and Roksana Boreli. *Is More Always Merrier?: A Deep Dive into Online Social Footprints.* In Proceedings of the 2012 ACM Workshop on Workshop on Online Social Networks, WOSN12, pages 6772, New York, NY, USA, 2012. ACM.

[29] D. Correa, A. Sureka, and R. Sethi. *WhACKY! - What Anyone Could Know about You from Twitter.* In Proceedings of the 10th Annual International Conference on Privacy, Security and Trust, 2012, PST12, pages 4350, 2012.

# Appendix A

# Questionnaires

## A.1 Pre-Study Questionnaire

**Category I: Demographics**

Note: Provide an option for not to disclose.

(1) What is your age ?

(2) What is your gender ?

- Male
- Female
- Other

(3) What is your highest degree (education) ?

- High School (10th Class)
- Senior Secondary (12th Class)
- Graduate
- Postgraduate
- PhD
- Other

(4) What is the discipline of your education ?

- Computer Science and related
- Science and related
- Commerce and related
- Humanities and related
- Others

(5) What is your occupation ?

- Student
- Faculty

- Service
- Private job
- Self-Employed
- Others

**Category II. Social Media Usage**

(1) Please mark the social networks on which you maintain your presence.

- Facebook
- Twitter
- Instagram
- LinkedIn
- Any Other (please specify)

(2) Where do you access your social media accounts ?

- Home
- Travel / Commuting
- Office / College
- Restaurants / Hangout places
- Any other

(3) How much time (approx) you spend on these social media accounts each day ?

- Less than 1 hour
- 1 to less than 3 hours
- 3 to less than 5 hours
- More than 5 hours
- Do not keep track

**Category III. Privacy within and across Social Media Accounts (User preferences, User behavior, Privacy perceptions and controls)**

(1) How do you prefer to use social networks in general ?

- Don't use social networks
- Use it only to get updates
- Use it to post rarely along with getting updates
- Use it to post frequently and obtain updates

(2) For which of the following purpose(s) do you use online social networks?

- Personal networking (family & relatives)
- Connect with professional colleagues & get updates on professional trends

- College friends and hangout updates
- Marketing purpose (self branding / promotion / opinion sharing)
- Any other (please specify)

(3) Who you think is the intended audience on following social networks that you use (tick all applicable)?

| Social Network | Family | Friends | Colleagues | Other |
|---|---|---|---|---|
| FB | | | | |
| TW | | | | |
| IN | | | | |
| LK | | | | |
| Any Other | | | | |

(4) Do you cross-post (post same content) across multiple social networks?

- Yes
    * What are the network pairs where you often cross post (tick all applicable) ?

| Network Pair | FB | TW | IN | LK |
|---|---|---|---|---|
| FB | | | | |
| TW | | | | |
| IN | | | | |
| LK | | | | |

- No

(5) While maintaining accounts on social media, what are your preferences with respect to user profile attributes ?

- Make sure all or most profile attributes are filled
- Don't bother too much, fill some profile attributes and leave others blank.
- Fill only the minimal (required) profile attribute values

(6) Answer the following in the context of 'user profile attribute values across different social networks' ?

(a) Do you have similar **username** across two or more social media accounts ?
- Yes
- No
    * Wanted similar username but it was already in use
    * Didn't bother too much

(b) Have you put up your **full name** across two or more social media accounts ?
- Yes
- No

(c) Have you entered your **location** across two or more social media accounts ?
- Yes

- No

(d) Have you entered your **profile pic** across two or more social media accounts ?

- Yes
- No

(7) Do you have common friends across your multiple social media accounts?

- Yes
- No

(8) How comfortable you are (on *likert scale*, 1: very uncomfortable, 2: uncomfortable, 3: not sure, 4: comfortable & 5: very comfortable) in each of the following scenarios ?

(a) If a user X (who is your friend) on **source** social network A can find your account on **target** social network B, will you be comfortable?

| Source Network (User X) | Target Network B | | | |
|---|---|---|---|---|
| | FB | TW | IN | LK |
| FB | | | | |
| TW | | | | |
| IN | | | | |
| LK | | | | |

(b) If a user X (who is your friend) on **source** social network A can see your activities your account on **target** social network B, will you be comfortable?

| Source Network (User X) | Target Network | | | |
|---|---|---|---|---|
| | FB | TW | IN | LK |
| FB | | | | |
| TW | | | | |
| IN | | | | |
| LK | | | | |

**Category IV. Awareness of Privacy Enabling Options**

(1) Are you aware of the privacy settings or controls in following social media platform ? (Answer on *likert scale*, 1: Not aware at all, 2: Unaware, 3: Not sure, 4: Aware & 5: Quite aware)

| Social Media | Awareness |
|---|---|
| FB | |
| TW | |
| IN | |
| LK | |
| Any Other | |

(2) Can you control the target audience of your activities on each of the following social media platform ? (Answer 1: Yes, 2: No & 3: Not sure)

| Social Media | Awareness |
|---|---|
| FB | |
| TW | |
| IN | |
| LK | |
| Any Other | |

(3) Have you ever tried to expose your activities to only specific targeted audience ? (Answer on *likert scale*, 1: Never, 2: Rarely, 3: Can't say, 4: Sometimes & 5: Mostly)

| Social Media | Awareness |
|---|---|
| FB | |
| TW | |
| IN | |
| LK | |
| Any Other | |

(4) Have you ever changed privacy settings of your profile on same social network ? (Answer on *likert scale*, 1: Never, 2: Rarely, 3: Can't say, 4: Sometimes & 5: Mostly)

| Social Media | Awareness |
|---|---|
| FB | |
| TW | |
| IN | |
| LK | |
| Any Other | |

(5) Have you ever changed privacy settings of your activity (say post) on same social network ? (Answer on *likert scale*, 1: Never, 2: Rarely, 3: Can't say, 4: Sometimes & 5: Mostly)

| Social Media | Awareness |
|---|---|
| FB | |
| TW | |
| IN | |
| LK | |
| Any Other | |

## A.2 Post-Study Questionnaire

### Category I. Basic Information

(1) What is your email ?

(2) Did you understand the broad concept of linkability score ?

- Not at all
- Little bit
- Most of it

- Completely

(3) Are you more aware (and more informed than you were earlier) about the linkability of your multiple identities across OSNs?

- Not at all, it knew it already
- Not really
- Somewhat yes
- Absolutely surely I am more aware now

(4) Did you notice the factors contributing to linkability scores on different pairs of OSNs?

- Did not notice at all
- Noticed it but didn't pay attention
- Noticed it carefully but don't remember the factors now
- Noticed and remember the factors as well

**Category II. Control Period**

(1) How many posts and profile changes you made during control period ?

- 0-2 posts, 0-2 profile changes
- 3-5 posts, 3-5 profile changes
- 6-8 posts, 6-8 profile changes
- $>$ 9 posts, $>$ 9 profile changes

(2) How many times you cross-posted (making same/similar post on two or more OSNs) during control period ?

- 0 - 2 times
- 3 - 5 times
- 6 - 8 times
- $>$ 9 times

**Category III. Treatment Period**

(1) How many posts and profile changes you made during treatment period ?

- 0-2 posts, 0-2 profile changes
- 3-5 posts, 3-5 profile changes
- 6-8 posts, 6-8 profile changes
- $>$ 9 posts, $>$ 9 profile changes

(2) How many times you cross-posted (making same/similar post on two or more OSNs) during treatment period ?

- 0 - 2 times

- 3 - 5 times
- 6 - 8 times
- $>$ 9 times

(3) How many times you received pop-up notification nudge during treatment period ?

- 0 - 2 times
- 3 - 5 times
- 6 - 8 times
- $>$ 9 times

(4) How many times did you change range of linkability score between any pair of OSNs during treatment period ?

- 0 - 2 times
- 3 - 5 times
- 6 - 8 times
- $>$ 9 times

(5) How many times did you change profile attributes in order to bring linkability scores within the desired range between any pair of OSNs during treatment period ?

- 0 - 2 times
- 3 - 5 times
- 6 - 8 times
- $>$ 9 times

(6) How many times did you modify or cancel a message draft to avoid red border around the message textbox on any OSN during treatment period ?

- 0 - 2 times
- 3 - 5 times
- 6 - 8 times
- $>$ 9 times

**Category IV. Conclusion**

(1) Which aspect/feature of nudge did you find most useful ? (Give your rating from 1 to 5 against each feature, 1: least useful and 5: most useful)

- Pop-up notification on top right
- Red color border while posting
- Value of linkability scores for each pair of OSNs
- Pie-charts depicting contribution of attributes in linkability scores

(2) What is your overall view of the linkability nudge ? (Give your rating from 1 to 5 against each view point, 1: least agree and 5: most agree)

- Not useful at all
- Useful but usage is complicated
- Useful and easy to use as well