Usage and Consequences of Privacy Settings in Microblogs

Usage and Consequences of Privacy Settings in Microblogs

Elena Daehnhardt, Nick K. Taylor
Department of Computer Science
Heriot-Watt University
Edinburgh, United Kingdom
Email: ed123@hw.ac.uk

Yanguo Jing
Faculty of Life Sciences and Computing
London Metropolitan University
London, United Kingdom
Email: y.jing@londonmet.ac.uk

***Abstract*—Twitter facilitates borderless communication, informing us about real-life events and news. To address privacy needs, Twitter provides various security settings. However, users with protected profiles are limited to their friendship circles and thus might have less visibility from outside of their networks. Previous research on privacy reveals information leakage and security threats in social networks despite of privacy protection enabled. In this context, could protecting microblogging content be counterproductive for individual users? Would microbloggers use Twitter more effectively when opening their content for everyone rather than protecting their profiles? Are user profile protection features necessary? We wanted to address this controversy by studying how microbloggers exploit privacy and geo-location setting controls. We followed a set of user profiles during half of year and compared their usage of Twitter features including status updates, favorites, being listed, adding friends and follower contacts. Our findings revealed that protecting user accounts is not always detrimental to exploiting the main microblogging features. Additionally, we found that users across geographic regions have different privacy preferences. Our results enable us to get insights into privacy issues in microblogs, underlining the need of respecting user privacy in microblogs. We suggest to further research user privacy controls usage in order to understand user goals and motivations for sharing and disclosing their microblogging data online with the focus on user cultural origins.**

***Keywords*-Twitter microblogs; privacy online; user content protection; geo-location usage; microblogging behavior;**

I. Introduction

Internet data often include cues on user online behavior and personal information, collected and processed by web services and applications. Social web sites such as Twitter and related web/mobile applications allow their users to connect with friends, sharing information in real time. Moreover, microblogs and other online resources can be exploited in business settings for marketing and research purposes. However, the microbloggers' personal privacy needs to be observed and weighted against the practical benefits provided by microblogs.

Online users' privacy could be supported with help of a regulatory framework and software controls. In support of the human right for privacy, national and international regulations are being developed to preserve personal privacy in an online setting. In this respect, the Organization for Economic Co-operation and Development guidelines [1] address online privacy protection and safe information transfer via computer networks, to prevent unlawful personal data access, storage and processing. In support of privacy and for their users' benefit, Social Networking (SN) web sites and applications provide functionality for users to exercise privacy control of shared user-generated content and metadata. For this, different user profile settings and options across SN web sites help users to hide sensitive information. Software means and privacy settings are however not yet very effective. Personal information can still be derived out of social networks with information retrieval, data mining and machine learning methods [2], and named-entity recognition techniques [3], [4]. Also, privacy of sensitive microposts can be violated by the user's friends reposting originally protected content [5]. Some Twitter software clients facilitate information leakage from protected users [5]. Sensitive topics on diseases or alcohol consumption could also be revealed by Twitter users, which could benefit from software tools protecting users from posting sensitive messages [4].

Despite of the "openness" culture widely appreciated in Twitter, we argue that for some of the users privacy still matters. Due to the lack of information on privacy issues and insufficiency of the privacy protecting mechanisms, it is paramount to further investigate real user needs, which might be governed by different purposes and modes of microblogging usage. For this, we overview the privacy protection means in Twitter. We analyze user behavior online to rise awareness towards protecting privacy in Twitter. While observing how privacy settings are exploited by Twitter users, we focus our analysis on the usage of Twitter profile protection and geographic location sharing features. We followed a set of users for a period of about six months and analyzed their privacy controls' usage with Twitter Application Program Interface (API). We distinguish between different usage purposes, which we relate with the usage of geo-location and profile protecting features. Our main contribution is to study online privacy controls usage for a set of selected user groups in Twitter. Next, we outline the main issues and means of protecting user privacy online with a focus on SN and microblogs.

II. Related Work

Twitter would be much less useful for sharing news and information in real-time and finding users with similar interests online, if everyone on Twitter would protect their status updates. Protecting Twitter messages might be seen as counterproductive for reaping business opportunities online. However, in some cases users opt to protect their tweets to safeguard their personal data. In this section, we discuss the privacy protection on Twitter and other social networking applications, possible threats and solutions.

Transparency and User Control: Social networking web sites such as Facebook, Twitter and LinkedIn often maintain user data including a user's network connections, geographic locations and employment positions respectively, when provided by the user. Some web applications collect user personal details explicitly or collect their online activities with user permission [6]. However, some of the web sites do not inform their users about data collection performed [6]. To ensure that user personal details are not shared unwillingly, users should exercise complete control over their personal details. The consequences of sharing the personal data should be thoroughly thought by the users, which ideally should have a complete control over their data [7]. [8] suggests to exploit personal data stores, in which users manage their privacy settings while remaining sole owners of their data. Service providers could further adjust services to available user data and associated privacy settings [8].

Mining Personal User Data: Furthermore, privacy is a paramount factor influencing microblogs' adoption by users [9]. Microblogs might be perceived as unsafe compared to other social platforms and therefore not exploited by some of the users, who are quite cautious about sharing their personal information online [9]. Indeed, social web makes it possible to mine openly available user data [7]. Users might publicly share their preferences, or user activities and traits could be automatically mined based on their online behavior [7]. In result, user behavior patterns could be exploited to deduce user specific traits. For instance, user geographic location could be inferred based on user-generated content, metadata associated with microblogs and user social network [2], [10]. Personal information can also be derived out of user social network, online web resources with help of data mining and named entity recognition techniques [3]. Thus, explicit user personal information sharing is not required to collect potentially sensitive information [7].

Information Leaks: Twitter enable users to protect their accounts. However, due to the nature of microblogs and their attractive openness, only a small fraction of about 10% of users protected their accounts [11]. It is not trivial to manage private information amongst friendship networks [12]. Indeed, even protected tweets can become publicly visible when being reposted by the user's friends [5]. In [5] dataset, about of 1% of all accounts retweeted private information of their contacts. The study into how the personal information can be revealed in microblogs by [13], exploiting machine learning and human annotations to estimate the level of sharing of personal information by microbloggers. As [12] pointed out, existing privacy issues do not stop online users willing to communicate with their friends and acquaintances. A "privacy score" could be applied to user contacts to help in decision making on how much information could be shared with these contacts [13].

Furthermore, as one of the solutions for supporting user privacy [14] suggest to use anonymizer methods to avoid personal information leakage. A prototype using cryptography and access control of followers was proposed by [15] in order to protect user-generated content including hashtags. In support of personal data protection, various methods of information coding were developed such as [16]. However, the coded information may be detrimental to its usage or affect the quality of the stored information [16]. This might be inconvenient for using coded information in practical applications, since SN web sites are specially build to facilitate users communication and sharing content. Thus, usage of these applications and services requires users to share information to a certain extend, enabling them to exploit a web site functionality as they require.

Privacy and Cultures: When referring to the human right regulations, [17] discusses perceptions of privacy in different cultural settings. It seems that societal values impact the view of privacy and regulations for different regions, for instance, in European countries and the United States [17]. In [4] authors found differences in the information leaks, particularly on the topic of "depression", for Singapore, the UK, and the US. The level and type of personal information sharing differs for users from the USA and Singapore, which tend to reveal personal information (contact, demographic, education and job) or their feelings respectively [18]. It is reasonable to assume that the privacy settings could be used differently amongst cultures. This is why it is interesting to further investigate how the privacy controls are exploited by different cultural groups in Twitter.

To summarize, Twitter microblogs allow users to share their messages and interesting links. The majority of Twitter accounts are open. Protected accounts are prone to the information leakage and require further care when sensitive or personal information is shared within friendship networks. It seems however that protecting tweets might be counterproductive and unnecessary, since protected profiles might block users from new personal or business opportunities in detriment of fostering user communication on Twitter. We further analyze whether protected users communicate and form relations less intensively compared with the open users and which in result might take more advantage of microblogging. Finally, we are going to investigate whether different privacy perceptions across the cultural regions could affect privacy controls usage.

Table I
FEATURES ANALYSED

Feature	Description
INFLUENCE	Ratio of Followers to Following (Friends)
STATUSES	Number of published tweets
FAVOURITES	Number of favorites user posted
LISTED	Number of lists in which user was included
FOLLOWERS	Number of followers
FRIENDS	Number of friends
SOURCES	Number of Twitter applications exploited
CHANGES	Number of setting changes (enabling/disabling geo-location services or protecting/opening user profiles)

Table II
CULTURAL DIMENSIONS AND ASSOCIATED COUNTRIES (WE JOINED COUNTRIES IN RELATION TO THEIR PROXIMITY TO THE APEXES (LA, MA, RE) OF TRIANGLE REPRESENTING THE LEWIS MODEL)

Cultural Dimensions	Top Countries
Linear-Active (LA)	The United States (US) and Great Britain (GB)
Multi-Active (MA)	Russia (RU), Brazil (BR), Spain (ES), Turkey (TR), France (FR), Mexico (MX), Italy (IT)
Reactive (RE)	Japan (JP) and Indonesia (ID)

III. METHODOLOGY

The main research question was to find out if protecting user accounts hampers an effective communication in Twitter. We selected a non-exhaustive list of features mostly available in the Twitter profile and listed in Table I. Next, we compared protected users activity with microblogging activity of users with open profiles, and tested the hypothesis:

- H1. Number of friends (FRIENDS feature) which user follows is fewer for protected user accounts;
- H2. Number of followers (FOLLOWERS) is smaller for the protected user;
- H3. User influence (INFLUENCE), defined as ratio of followers to friends, is smaller for the protected users;
- H4. Status updates (STATUSES), or twitter microblogs posted by the user, are less posted by protected users;
- H5. Number of lists (LISTED) in which user was included is smaller for the protected users;
- H6. Number of favorites (FAVOURITES) is smaller for protected users.

Since privacy control designs differ amongst Twitter software clients, we also analyze usage of Twitter clients. Our next hypotheses assume that protected users tend to exploit and try out different software products and setting changes in the beginning of Twitter usage, in order to find out which software and settings fits the best to their needs:

- H7. Number of setting changes (CHANGES) of protected users is greater compared with the "open users".
- H8. Number of software clients (SOURCES) of protected users is greater compared with the "open users".

In order to detect user cultural preferences for privacy settings, we adopt the cultural dimensions from the Lewis Model of Cultures [19] to establish cultural profiles for the selected countries in Table II. Based on the trait of showing or concealing personal feelings determining cultural differences, as defined in the Lewis questionnaire[1], we assumed that Multi-Active (MA) users such as originating from countries like Russia, Brazil or Spain are likely to open their user profiles and communicate their opinions and feelings for a greater audience. In contrast, Reactive (RE) users from Japan and Indonesia tend to conceal their feelings, which might be reflected in their preference to close or protect their Twitter accounts. Linear-active (LA) users from the United States of America and Great Britain prefer to partly not expose their feelings and therefore they might have a smaller fraction of protected profiles compared to RE, and smaller fraction of open profiles as compared to MA users. This is why we established the following hypothesis:

- H9. MA users prefer open profiles the most.
- H10. RE users prefer closed profiles the most.

Experimental Setup: It is important to mention that users change their settings over time and we analyze which settings are mostly exploited by each particular user. This is why we follow users who started their microblogs on the same day and observe how their settings change for about half of year as follows:

- STEP 1: Collect about 21600 users registered with Twitter on 26th November 2014 by listening to the Twitter sample stream for a period of about three days, from 26th to 29th November 2014;
- STEP 2: Visit the selected user profiles in a period of about six months and monitor their usage of the geographic location sharing and profile protection features, statistics on status updates and friendship/followers network growth in time; It is important to mention that Twitter web application and API enable us to observe aforementioned features (such as number of connections and status updates) out of protected user profiles, therefore, our experimental setup does not violate user privacy while working with aggregated usage statistics.
- STEP 3: In order to understand user needs for protecting their microblogging content in respect of their cultural origin, we exploited our country-detecting Multinomial Naive Bayes classifier built on text features extracted from user profile meta-data and joining free-text of user-defined location, preferred language and time zone, with a three-times cross-validation accuracy of about 90%, as described in [2].
- STEP 4: Analyze and interpret the data collected for addressing the hypothesis stated above. For finding the privacy needs of the users from respective cultural

[1] http://bestcareermatch.com/cross-cultural-communication

Table III
PRIVACY AND GEO-LOCATION SETTINGS IN " PRIVACY AGGREGATED" DATASET: T (TRUE) FOR ENABLED SETTING, OTHERWISE F (FALSE)

Geo-location	Protected	Verified	Description	Number of Users
F	F	F	Open profiles without geo-location setting activated	17182
T	F	F	Open profiles with geo-location setting enabled	2973
F	T	F	Protected profiles without geo-location	733
T	T	F	Protected profiles with geo-location setting activated	243
F	F	T	Protected profiles of public persons (or celebrities) disabled geo-location setting	2

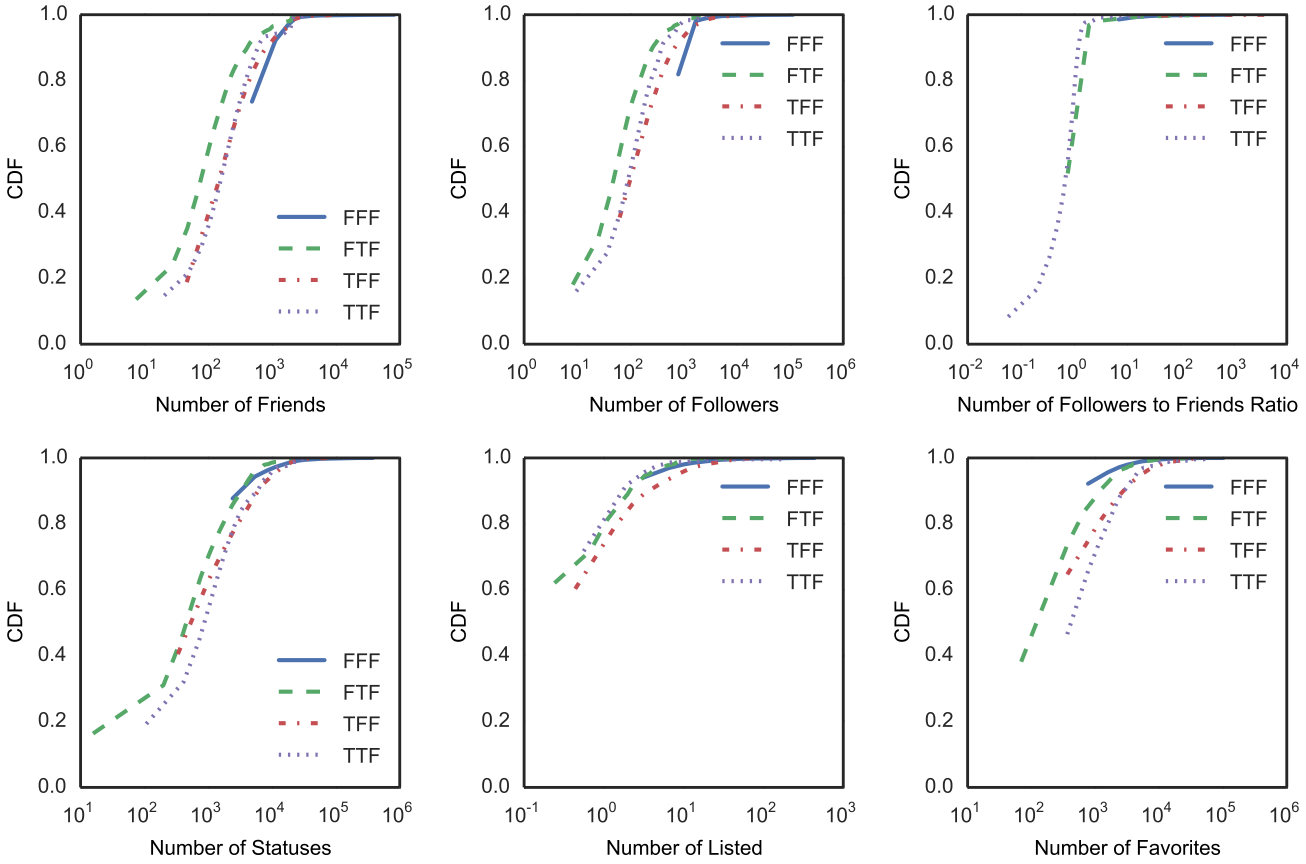

Figure 1. Networking and Twitter Features Usage with Different Profile Settings

regions, we grouped countries into cultural dimensions (regions) as defined by the Lewis model of Cultures.

Data Collection: In the first step, we collected "Tweets" dataset consisting of 48287 tweets published by 22624 users from 26th to 30th November 2014. These are tweets published by users registered with Twitter on 26th of November 2014. Out of the set of users from "Tweets" dataset, we randomly selected 21600 users to be further followed for the next six months. We exploited our three Twitter accounts with Twitter Representational State Transfer (REST) API for retrieving user profiles of the selected users. These user profile settings were stored into "Privacy" dataset including information on the number of published tweets, followers, friends, listed, favorites count; and also on the geo-location settings turned on, verified (if person is a celebrity or verified on Twitter) and protected (their content is only visible to their friends) settings.

It is important to mention, that we do not consider daily change patterns due to Twitter REST API limitations[2] imposed. Due to server maintenance needs and while starting our data collection, we experienced some interruptions while storing user profile settings, which did not effect our final results based on the aggregated statistics for the whole data collection time.

We observed that some of the users change their settings to enable geo-location sharing, while others protect their profiles. We also observed an increase of protected profiles, from 2.4% to 7.3%, and geo-location sharing profiles, from

[2]https://dev.twitter.com/rest/public/rate-limits

7.5% to 19%, towards the end of data collection.Having the data on Twitter usage and user settings for the sampled half of year, we aggregated the most exploited settings of client software usage, geo-location sharing and profile protection, and the maximal number of tweets, friendships, lists and favorites. These data comprised a "Privacy Aggregated" dataset, which we further study to find out general behavior differences amongst Twitter users having certain privacy and geo-location sharing preferences.

IV. Findings

To summarize for the whole data collection period, about 95% of our users prefer to keep their user profiles openly available. Geographical location sharing feature is exploited mostly by 15% of our users. Table III above lists our main setting combinations. It is important to mention that the FFT group included "verified" users of two public persons, in our case a music band and politian, which are disregarded in the group comparisons when performing unequal variances statistical tests (Table IV). We observed that these two users have larger social networks when compared with other user groups. They also included into more listings. The greatest group of 17182 users (FFF) prefers open profiles without geographic information available. The second largest group of 2973 users (TFF) prefers open profiles while sharing their geo-locations.

Friends and Followers : Interestingly, the cumulative distribution function with logarithmic scale helps to visualize differences between user groups (Figure 1 above). Particularly, protected users with enabled geo-location services (TTF) have a greater number of FRIENDS and FOLLOWERS compared with users with closed profiles without geo-location settings (FTF). The INFLUENCE of the latest is however not significantly different from the former ones (Table IV).

Status updates, Listed and Favorite Counts : Users (TTF and TFF) with protected and open profiles, with geo-location services enabled do post more STATUSES in average compared to FTF and FFF respectively. TTF and FFF has no significant differences in their status updates (Table IV). Users with open and geo-enabled profiles (TFF) tend to be included into more lists (LISTED), while the protected users with geo-sharing settings (TTF) compete in the number of FAVOURITES with open users who do share their geo-locations (TFF) as well.

Frequency of Setting Changes and Software Clients: Table IV shows statistically significant differences in CHANGES feature between open and closed user profiles. Geo-location sharing users (TFF) changed their applications (SOURCES) more often compared with the users with protected accounts (FTF), which in turn change their settings more in average. Despite of user profile protection, Twitter still provides information on client software usage shown in Table V below. Twitter for Android software is used

Table IV
Unequal variances t-test for various Settings: Mean (μ), Standard Deviation (σ), Welchs Test statistic (t), two-tailed p-value (p)

Feature	μ_{gr_1}	σ_{gr_1}	μ_{gr_2}	σ_{gr_2}	t	p
$Group_1$ ($N_{gr_1} = 17182$): FFF vs. $Group_2$ ($N_{gr_2} = 733$): TTF						
INFLUENCE	2.72	105.46	1.76	19.31	0.88	0.37
STATUSES	1810.87	7844.50	1208.69	2321.30	5.75	< 0.01
FAVOURITES	323.78	1793.38	503.16	1782.19	-2.66	< 0.01
LISTED	1.27	7.37	0.74	2.53	4.85	< 0.01
FOLLOWERS	367.30	1902.00	119.79	245.46	14.46	< 0.01
FRIENDS	421.12	1591.38	183.78	349.27	13.39	< 0.01
SOURCES	1.03	0.20	1.03	0.18	0.26	0.78
CHANGES	1.08	0.32	1.92	0.39	-56.34	< 0.01
$Group_1$ ($N_{gr_1} = 17182$): FFF vs. $Group_2$ ($N_{gr_2} = 2973$): TFF						
INFLUENCE	2.72	105.46	2.39	68.48	0.21	0.82
STATUSES	1810.87	7844.50	2144.56	4579.74	-3.23	< 0.01
FAVOURITES	323.78	1793.38	1128.26	3391.42	-12.63	< 0.01
LISTED	1.27	7.37	1.81	6.14	-4.27	< 0.01
FOLLOWERS	367.30	1902.00	320.75	978.62	2.01	< 0.05
FRIENDS	421.12	1591.38	339.87	632.67	4.83	< 0.01
SOURCES	1.03	0.20	1.05	0.23	-3.39	< 0.01
CHANGES	1.08	0.32	1.68	0.69	-45.79	< 0.01
$Group_1$ ($N_{gr_1} = 17182$): FFF vs. $Group_2$ ($N_{gr_2} = 243$): TTF						
INFLUENCE	2.72	105.46	0.75	1.18	2.43	< 0.05
STATUSES	1810.87	7844.50	2157.59	4259.59	-1.23	0.21
FAVOURITES	323.78	1793.38	1301.08	4150.01	-3.66	< 0.01
LISTED	1.27	7.37	1.27	10.93	0.00	0.99
FOLLOWERS	367.30	1902.00	182.81	349.50	6.90	< 0.01
FRIENDS	421.12	1591.38	298.31	474.33	3.74	< 0.01
SOURCES	1.03	0.20	1.03	0.17	0.26	0.79
CHANGES	1.08	0.32	2.76	0.74	-35.13	< 0.01
$Group_1$ ($N_{gr_1} = 733$): FTF vs. $Group_2$ ($N_{gr_2} = 2973$): TTF						
INFLUENCE	1.76	19.31	2.39	68.48	-0.43	0.66
STATUSES	1208.69	2321.30	2144.56	4579.74	-7.79	< 0.01
FAVOURITES	503.16	1782.19	1128.26	3391.42	-6.90	< 0.01
LISTED	0.74	2.53	1.81	6.14	-7.29	< 0.01
FOLLOWERS	119.79	245.46	320.75	978.62	-9.99	< 0.01
FRIENDS	183.78	349.27	339.87	632.67	-8.99	< 0.01
SOURCES	1.03	0.18	1.05	0.23	-2.18	< 0.05
CHANGES	1.92	0.39	1.68	0.69	12.60	< 0.01
$Group_1$ ($N_{gr_1} = 733$): FTF vs. $Group_2$ ($N_{gr_2} = 243$): TTF						
INFLUENCE	1.76	19.31	0.75	1.18	1.41	0.15
STATUSES	1208.69	2321.30	2157.59	4259.59	-3.31	< 0.01
FAVOURITES	503.16	1782.19	1301.08	4150.01	-2.90	< 0.01
LISTED	0.74	2.53	1.27	10.93	-0.74	0.45
FOLLOWERS	119.79	245.46	182.81	349.50	-2.60	< 0.01
FRIENDS	183.78	349.27	298.31	474.33	-3.46	< 0.01
SOURCES	1.03	0.18	1.03	0.17	0.08	0.92
CHANGES	1.92	0.39	2.76	0.74	-16.85	< 0.01
$Group_1$ ($N_{gr_1} = 2973$): TFF vs. $Group_2$ ($N_{gr_2} = 243$): TTF						
INFLUENCE	2.39	68.48	0.75	1.18	1.30	0.19
STATUSES	2144.56	4579.74	2157.59	4259.59	-0.04	0.96
FAVOURITES	1128.26	3391.42	1301.08	4150.01	-0.63	0.52
LISTED	1.81	6.14	1.27	10.93	0.76	0.44
FOLLOWERS	320.75	978.62	182.81	349.50	4.80	< 0.01
FRIENDS	339.87	632.67	298.31	474.33	1.27	0.20
SOURCES	1.05	0.23	1.03	0.17	1.51	0.13
CHANGES	1.68	0.69	2.76	0.74	-21.98	< 0.01
$Group_1$ ($N_{gr_1} = 20155$): Open FFF and TFF vs. $Group_2$ ($N_{gr_2} = 976$): Protected FTF and TTF						
INFLUENCE	2.67	100.86	1.51	16.75	1.30	0.19
STATUSES	1860.09	7454.24	1444.94	2952.54	3.83	< 0.01
FAVOURITES	442.45	2125.85	701.82	2603.56	-3.06	< 0.01
LISTED	1.35	7.20	0.87	5.87	2.45	< 0.05
FOLLOWERS	360.44	1795.96	135.48	276.22	14.57	< 0.01
FRIENDS	409.13	1489.56	212.30	387.15	12.12	< 0.01
SOURCES	1.03	0.21	1.03	0.18	0.74	0.45
CHANGES	1.17	0.45	2.13	0.62	-47.51	< 0.01

Table V
TOP TWITTER CLIENTS: SOFTWARE (PERCENT OF USERS)

Client	FFF	TFF	FTF	TTF
Twitter for Android	24%	52%	38%	51%
Twitter for IPhone	23%	26%	40%	34%
Twitter Web Client	14%	13%	9%	6%

Table VI
PERCENT OF USERS BY CULTURE AND COUNTRY-GROUPS

Culture	Country Codes (% of users)
LA	US (43.4), GB (1.0)
MA	RU (12.0), BR (8.6), ES (7.5), TR (6.6), FR, MX & IT (2.5)
RE	JP (16.1), ID (2.4)

by about 30% of our users, followed by Twitter for IPhone used by 24% and Twitter Web Client used by almost 14% of users. Twitter Web Client is used the least by geo-enabled users with closed profiles (TTF). It seems that this user group prefers to use mobile device applications instead.

Cultural Differences in Privacy Settings Usage: Table VI summarizes users classified into their country and cultural groups, which are compared according to their profile protection settings. Since we exploited country and culture predictive models developed in [2], we further assessed their performance on a set of 145 users having countries referred in their geo-enabled tweets. We observed that human labels of the 145 user locations matched with the defined geo-location provided by Twitter in about 86% of cases. We achieved about 83% and 94% accuracy, 78% and 94% precision, 83% and 94% recall for the country and culture group classifications, respectively while comparing the classification output with the test set of 145 users. Overall, our cultural group classifier enabled to predict 85% (40 out 47 users) of LA, 97% of MA (73 out of 75 users) and 100% of RE (23) users.

Further, we exploited the culture group classification model to analyze usage of protected and open profiles for our users set. We found out that MA users have a larger percentage (98%) of open profiles, while RE have the greatest percentage (11%) of closed profiles compared to other cultural groups. Therefore, we could accept our hypotheses H9 and H10. It seems, that privacy perceptions or needs differ amongst the cultural groups analyzed.

Summary: Our sample distributions could not satisfy normality and equal variances assumptions in the majority of cases. This is why we selected Welch's unequal variances statistical test (using Scipy Python library) to compare users within particular setting groups. This test allowed us to test if two independent samples have similar averages. When p-values were $p < 0.05$, we disregarded the null hypothesis that the averages are similar. Table IV shows results for comparing means between paired setting groups, which we used to accept or reject our research hypothesis, revisited in Table VII. We arranged Table IV to compare settings

Table VII
HYPOTHESIS REVISITED: WHILE COMPARING OPEN (FFF+TFF) AND PROTECTED (FTF+TTF) USER PROFILES

	Hypothesis	Conclusion
H1	Protected users have fewer friends	Accepted
H2	Protected users have fewer followers	Accepted
H3	Protected users are less "influential"	Rejected
H4	Protected users have fewer status updates	Accepted
H5	Protected users are less listed	Accepted
H6	Protected users have less favorites	Rejected
H7	Number of setting changes of protected users is greater compared with the open users	Accepted
H8	Number of software products (SOURCES) of protected users is greater compared with the open users	Rejected
H9	MA users prefer open profiles the most	Accepted
H10	RE users prefer closed profiles the most	Accepted

groups for open profiles, including FFF (open profiles without geo-location services enabled) and TFF (open profiles with geo-location enabled), and protected profiles, including FTF (closed profiles) and TTF (closed profiles with geo-location services enabled). At the bottom of the table we placed the feature comparison for answering our hypothesis statements for merged user groups, with open (FFF and TFF) and protected (FTF and TTF) profiles.

Overall, we accept H4 and H5, since users with publicly open profiles exploit the Twitter features STATUSES and LISTED the most compared with users with protected user profiles. Generally, open profiles attract more followers and have more friends, such that we could accept our hypotheses H1 and H2. However, we could not find significant differences in the INFLUENCE feature for the geo-location enabled users (TFF and TTF). As seen from the bottom of the Table IV, the INFLUENCE does not differ significantly for users with open and protected profiles. Similarly, the number of SOURCES is comparable for both user groups. This is why we cannot accept hypotheses H3 and H8.

Interestingly, users with protected profiles (FTF and TTF) tend to exploit the FAVOURITES feature more actively in contrast with open profile users (FFF). However, when users with open profiles enabled their geo-location settings (TFF), their favoring statistics was not significantly different compared with the users with closed profiles (TTF). When comparing overall PROTECTED and OPEN user groups, we found that PROTECTED user profiles exploit more FAVOURITES compared to OPEN user profiles, and therefore rejecting H6. We accepted H7 since protected users performed more setting changes as compared to users with open profiles.

Furthermore, the protected user profiles with the geo-location feature enabled (TTF) showed no significant differences when compared with open user profiles (FFF and TFF) in the number of STATUSES, LISTED and SOURCES. Thus, these users are quite active in their publishing behavior, since they are included in lists and exploit various devices and software clients. It seems, that we still need

to consider a profile protecting feature to address this user group needs. Their motivations of microblogging usage could further be investigated with a user feedback.

V. Discussion and Further Work

Reflection on Our Results: Even though protecting Twitter profiles might seem to be counterproductive due to the microblogging nature of networking, around 5% of our users prefer to close their accounts from public view. We observed an increase in the number of protected accounts about threefold in the half year period of following user accounts. Protecting user accounts in microblogs could however mislead users into a wrongly perceived safety, since personal data could still be automatically mined or revealed by online friends.

Moreover, no significant differences in user influence and number of exploited Twitter client applications for protected and open user profiles were found. However, the number of status updates, listed, followers and friends were greater for the open user profiles. The open user profiles with geo-location enabled were the most active user group in terms of all features we analyzed. One of the interesting findings is that protected user profile tend to favorite the most. Does it mean that they like to keep the favorite tweet for later and do not want to further propagate the tweet as when using retweet? Alternatively, favoring might mean that protected users personally appreciate authors of their favorite tweets and might motivate their following behavior. Additionally, we could not find significant differences of posting status updates and the number of list inclusions between geo-location enabled users with protected profiles and users with open profiles. We found also different fractions of open and closed user profiles for users across cultural regions. We think that further investigation into The purpose of different microblogging usage modes and privacy preferences, preferable with feedback of microbloggers actively using Twitter services and in respect of their cultural origins could be further investigated.

Privacy in Social Networks: Furthermore, more than a third part of user-generated content is shared with default privacy setting, which often does not correspond with user needs as revealed in a Facebook user survey [20]. As was discussed in [21], Facebook applications design might affect the usage of privacy-related features [15]. In Twitter, users can protect their content from public view, but online friends could reveal the initially protected information [5]. In result, anyone and with any intention could get access to private and sensitive user information, whether explicitly provided by users in openly available blogs, or mined with help of text analysis and other automated tools taking advantage of existing data mining and machine learning techniques. In real-life settings and considering users' willingness to communicate and social networks, which can leak one's private information, there is no really serious protection yet to ensure human privacy in microblogs, which could be further enhanced using cryptographic techniques, as in [15]. Additional security measures and further privacy regulations should be implemented, since the dangers of human privacy violations in microblogs cannot be disregarded.

Open vs. Commercial Exploitation: Twitter can also be used by commercial companies as a marketing and communication tool for influencing their customers, as studied in [22]. The effectiveness of Twitter and Facebook usage for hotel marketing purposes was investigated in [23], showing a strong relationship between social media user experience, hotel perceptions and further booking intentions. In opposite to openly available social media content, followers' control could open a possibility to paid services [15]. However, this approach could change microblogging as it is now, and its impact on openness of content and society at large will require further consideration.

Research Considerations: Twitter corporation provided their public tweets archive to the Library of Congress [24], [25]. Twitter states, that tweets collection opens new perspectives for research and ways to retrieve information related to past events [26]. Twitter together with Facebook and Buzz also made their public content searchable via Google Search engine [26]. Nevertheless, openly available Twitter content and meta-data could provide scientists with much required data for performing research experiments. However, would it be ethical to access and retain user data without appropriate consent and potentially infringing human privacy? The ethical dilemma of using Twitter data in research while protecting user privacy was discussed in [27], suggesting to gain institutional review boards approval before performing data collection of particular users. In our research, we deal mostly with aggregated user profiles while analyzing user behavior patterns and preferences. We avoid data retention of individual users, and anonymize user names and tweets when needed.

Further Work: Based on the previous works analyzing differences of sensitive information leaks [4] and personal and emotional information sharing in Twitter [18] for several countries, we assumed that the need for privacy control settings could be related to user cultural origins. Our findings reveal that persons from different cultural origins have their own preferences towards Twitter profiles protection. In further work, we aim to investigate the microblogging purposes and their relation to the usage of privacy mechanisms in view of user cultural context. For this, we plan to update and re-evaluate our classification models predicting user origins. Our evaluation could include larger training and test sets, and human evaluation of the classifiers.

VI. Conclusion

Above, we analyzed usage of Twitter profile protection and geo-location sharing controls. We were interested to find out if users with protected profiles do not exploit Twitter

features to their full advantage. For this, we statistically compared user groups of protected and openly available profiles in terms of their status updates, contact networks and other features. We found out that protected users have smaller social networks, which are however not less influential. Protected users actively favorite other content compared to the open-profile users without geo-location services enabled. When users with protected profiles enable their geo-location services, their tweeting behavior does not differ significantly from users with open profiles but without geo-location enabled. When users with open accounts enable geo-location services, they become the most active while using the aforementioned Twitter features, except of privacy setting changes. Protected users change their software settings the most. It seems, that users preferring to exploit protected profiles have their own motivation to microblogging. We observed cultural preferences in user profile protection settings usage, which can be further exploited for adapting web applications to specific cultural user profiles. Moreover, despite of the small fraction of protected user profiles, the human need for privacy in microblogs cannot be underestimated. We suggest more thorough exploitation of user generated content, while rising user attention towards possible privacy threats in microblogs. In further work, we plan to focus on cultural privacy perception differences in microblogs while extending our experimental setup with our previously developed user location detection classifier, which was re-evaluated with a new test set. In a nutshell, we suggest to analyze cultural groups' preferences rather than individual accounts to preserve human privacy and with a focus on satisfying user cultural preferences and needs.

REFERENCES

[1] OECD. (2013) The oecd privacy framework. [Online]. Available: http://www.oecd.org/sti/ieconomy/2013-oecd-privacy-guidelines.pdf

[2] E. Daehnhardt, Y. Jing, and N. Taylor, "Cultural and geolocation aspects of communication in twitter," in *ASE Internation Conference on Social Informatics 2014*, 2014.

[3] L. Zhang and W. Zhang, "An information extraction attack against on-line social networks," in *Social Informatics (SocialInformatics), 2012 International Conference on*. IEEE, 2012, pp. 49–55.

[4] H. Mao, X. Shuai, and A. Kapadia, "Loose tweets: an analysis of privacy leaks on twitter," in *Proc. 10th annual ACM workshop on Privacy in the electronic society*. ACM, 2011, pp. 1–12.

[5] B. Meeder, J. Tam, P. G. Kelley, and L. F. Cranor, "Rt@ iwantprivacy: Widespread violation of privacy settings in the twitter social network," in *Proc. The Web*, vol. 2, 2010.

[6] P. M. Schwartz, "Property, privacy, and personal data," *Harvard Law Review*, pp. 2056–2128, 2004.

[7] M. Hildebrandt, K. OHara, and M. Waidner, *Digital enlightenment yearbook 2013: The value of personal data*. IOS Press, 2013.

[8] E. Papadopoulou, A. Stobart, N. Taylor, and H. Williams, "Enabling data subjects to remain data owners," in *Agent and Multi-Agent Systems - Technology and Applications*, ACM. Berlin: Springer Verlag, 2015, pp. xx–xx.

[9] D. H.-L. Goh and A. Y. Chua, "Understanding the barriers to using microblogs," in *Proc. World Congress on Engineering and Computer Science*, vol. 1, 2013.

[10] E. Ilina (Daehnhardt), "A user modeling oriented analysis of cultural backgrounds in microblogging," *Human Journal*, vol. 1, no. 4, pp. 166–181, 2012.

[11] B. Rieder, "The refraction chamber: Twitter as sphere and network," *First Monday*, vol. 17, no. 11, 2012.

[12] J. Grimmelmann, "Saving facebook," *Iowa L. Rev.*, vol. 94, p. 1137, 2008.

[13] A. Caliskan Islam, J. Walsh, and R. Greenstadt, "Privacy detective: Detecting private information and collective privacy behavior in a large social network," in *Proc. 13th Workshop on Privacy in the Electronic Society*. ACM, 2014, pp. 35–46.

[14] E. McCallister, T. Grance, and K. Scarfone. (2010) Guide to protecting the confidentiality of personally identifiable information (pii): Recommendations of the national institute of standards and technology. [Online]. Available: http://csrc.nist.gov/publications/nistpubs/800-122/sp800-122.pdf

[15] E. De Cristofaro, C. Soriente, G. Tsudik, and A. Williams, "Hummingbird: Privacy at the time of twitter," in *Security and Privacy (SP), 2012 IEEE Symposium on*. IEEE, 2012, pp. 285–299.

[16] R. J. Bayardo and R. Agrawal, "Data privacy through optimal k-anonymization," in *Proc. 21st International Conference on Data Engineering, ICDE 2005*. IEEE, 2005, pp. 217–228.

[17] J. Q. Whitman, "The two western cultures of privacy: Dignity versus liberty," *Yale Law Journal*, pp. 1151–1221, 2004.

[18] W. Dong, M. Qiu, and F. Zhu, "Who am i on twitter?: A cross-country comparison," in *Proc. Companion publication of the 23rd International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 2014, pp. 253–254.

[19] R. Lewis, *When cultures collide: Managing successfully across cultures*. Nicholas Brealey Publishing, 2000.

[20] Y. Liu, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, "Analyzing facebook privacy settings: user expectations vs. reality," in *Proc. ACM SIGCOMM conference on Internet measurement conference*. ACM, 2011, pp. 61–70.

[21] M.-R. Ulbricht, "Privacy settings in online social networks as a conflict of interests."

[22] M. Bulearca and S. Bulearca, "Twitter: a viable marketing tool for smes," *Global Business and Management Research: An International Journal*, vol. 2, no. 4, pp. 296–309, 2010.

[23] X. Y. Leung, B. Bai, and K. A. Stahura, "The marketing effectiveness of social media in the hotel industry a comparison of facebook and twitter," *Journal of Hospitality & Tourism Research*, vol. 39, no. 2, pp. 147–169, 2015.

[24] A. Signorini, A. M. Segre, and P. M. Polgreen, "The use of twitter to track levels of disease activity and public concern in the us during the influenza a h1n1 pandemic," *PloS one*, vol. 6, no. 5, p. e19467, 2011.

[25] The Library of Congress. (2010) Twitter donates entire tweet archive to library of congress. [Online]. Available: http://www.loc.gov/today/pr/2010/10-081.html

[26] B. Stone. (2010) Tweet preservation. [Online]. Available: https://blog.twitter.com/2010/tweet-preservation

[27] C. M. Rivers and B. L. Lewis, "Ethical research standards in a world of big data [v1; ref status," *F1000Research 2014*, vol. 3, no. 38, 2014. [Online]. Available: http://f1000r.es/2wq