Post-print version of the article https://doi.org/10.1016/j.chb.2015.05.045

Comparing Twitter and Facebook User Behavior: Privacy and other Aspects

blind

Abstract

Understanding online-social-network (OSN) user behavior is an important challenge in the field of social network analysis, as OSNs play a significant role in people's daily lives. So far, many studies considering only one OSN or, at most, comparing results obtained for a single OSN, have been provided. Nowadays, users typically join more OSNs and this is an important aspect that should be taken into account for user behavior analysis. In this paper, we give an important contribution in this direction, by analyzing the behavior of users belonging to both `Facebook` and `Twitter`. This way, the analysis is well-founded because it is conducted on a common set of users and, further, a number of specific analyses become possible (as common friendship). Our study is carried out on data extracted from the web, and allows us to find important specificities of these users about their privacy setting, the choice of friends and the activity they do, which are generally consistent with the recent findings in this field.

Keywords: Social Network Analysis, `Twitter`, `Facebook`, Privacy.

1. Introduction

In a few years since their launch, online social networks (OSNs) have reached a huge level of popularity around the world. This rapid growth, leading OSNs to be large-scale, ubiquitous, and anytime services, has attracted the interest of researchers coming from disparate fields, also to study this new form of computer-mediated human interaction that facilitates people interaction and helps maintaining ties (Ellison et al., 2007). Understanding OSN user behavior is important to different entities (Jin et al., 2013): For Internet and OSN providers to guide infrastructural and application-level actions, for users themselves to enhance awareness in this potentially insecure world, for companies and government institutions to make better use of this huge network of people for their finalities, for scientists to better understand individuals and communities.

Several studies in the literature have analyzed many aspects of OSNs, such as connectivity, interaction, traffic activity, mobile social behavior, malicious behavior and privacy awareness (Maia et al., 2008; Cha et al., 2007; Gill et al., 2007; Pfeil et al., 2009; Hassan, 2009; Burke et al., 2009; Leskovec et al., 2007; Watts et al., 2007; Zhang et al., 2014b,a), and all the analyses have been conducted by focusing on only a single OSN (see Section 2). However, a multiple-social-network perspective may be much more fruitful to understand new aspects of people's interaction with OSNs. Indeed, each social network is a different environment providing a virtual "square" where a user expresses a different trait of personality (Barash et al., 2010), sometimes almost a different identity. The relevance of this perspective has been shown in the recent literature, from the point of view of both structural analysis (Buccafurri et al., 2015, 2014b, 2013) and applications (Zhang et al., 2015; Buccafurri et al., 2014a; Nguyen et al., 2013). However, less effort has been so far devoted towards comparative behavioral studies.

Observe that, in this case, the trivial comparison of the behavior of (different) users on different social networks does not give correct information, so we cannot just elaborate the results obtained in the literature in the different social networks. To give a trivial example, if we want to study how the behavior of a driver changes in cars A and B, we should study a sample of people driving both the cars, and observing the differences in the two experiences. We cannot simply study the expected behavior of drivers of car A and the expected behavior of drivers of car B and compare them, because the result would be affected by those traits prompted people to use car A instead car B. The same happens for comparative studies on behavioral aspects of online social networks, leading to the necessity of considering *membership overlap* (i.e., users belonging to all the studied OSNs) as the right perspective from which to draw meaningful and well-founded results.

Following the above multiple-social-network perspective, this paper aims to compare people's behavior in the two most popular social networks, which are `Facebook` and `Twitter`. On the basis of the previous observation, we base our analysis on the concept of membership overlap, to study a number of behavioral aspects.

The first one is about privacy and disclosure of personal information. Recent studies on `Facebook` have shown that both a strong association between low engagement and privacy concern (Staddon et al., 2012) and a significant relationship between privacy awareness and privacy concerns/self-disclosure (Zlatolas et al., 2015) exist. Our study aims to answer the question "Is there a connection between user awareness about privacy threats and membership overlap between `Twitter` and `Facebook`?".

The second aspect we study is about friendship. OSNs are important for maintaining social relations and previous studies have found that friendship is positively correlated with bridging social capital (Johnston et al., 2013; Burke et al., 2011; Bohn et al., 2014). As for this aspect, we study what is the attitude of users to have friendship relations overlapping between `Facebook` and `Twitter` and if a correlation between number of friends in `Twitter` and `Facebook` exists.

The last issue we deal with concerns the activity of users belonging to both `Twitter` and `Facebook`. Pempek et al. (2009) found that the prime goal of user activity on `Facebook` is to self-promote or to maintain relationships, whereas other studies showed that some types of activity are a sign of narcissism (Rosen et al., 2013; Carpenter, 2012; Bergman et al., 2011). Our study aims to answer the question "What about user activity and how the prevalence of activity on Facebook or Twitter is correlated to membership overlap?".

The answers to all these questions are interesting and sometimes surprising, showing that one of the contributions of this paper is to trace the beginning of a promising research line.

2. Related Work

With the increase in both the number and the dimension of OSNs, the development of approaches aiming to deeply investigate their main features has become welcome. A very interesting trend of the recent literature is to try to characterize user behavior in different platforms.

There are a number of relevant reasons that call for a deeply study on how users act when logging to these sites. First, analysis of user behavior allows the evaluation of the performance of existing social systems, more refined site design (Burke et al., 2009; Wilson et al., 2009) and user tailored advertisement placement policies (Williamson, 2007). Second, accurate models of user behavior in OSNs are crucial in applications of social studies such as viral marketing. Indeed, one of the main goals of marketers is the spread of their promotions quickly and widely. For this reason, they need to understand how users interact and to build models representing these interactions in such a way as to foresee how information will flow (Leskovec et al., 2007; Watts et al., 2007). Third, the study of user behavior helps the prediction on how much the future workload of OSNs will influence the whole Internet traffic, which is an essential information to properly dimension the Internet infrastructure and content distribution systems (Krishnamurthy, 2009; Rodriguez, 2009). Benevenuto et al. (2009) show an analysis on a clickstream dataset collected from a social network aggregator, providing users with a common interface for accessing multiple social networks.

As for the analysis about user social behavior in a social network, Teevan et al. (2011) and Java et al. (2007) explored search behavior on Twitter: while Teevan et al. (2011) make a deep analysis of large-scale query logs and supplemental qualitative data, Java et al. (2007) focus on the study of the topological and geographical properties.

Ross et al. (2009) examine a sample of undergraduate students to understand the nature of Facebook use. They study how personality and competency factors influence its use and how the Five-Factor Model of personality (openness, conscientiousness, extraversion, agreeableness, and neuroticism) is related to Facebook use.

A systematic measurement study on the statistics of the social network underlying the video sharing service YouTube is reported by Cheng et al. (2008), in which a deep analysis on user behavior through a number of ratings, views and comments on YouTube videos is carried out. Another study on YouTube is presented by Maia et al. (2008) and focuses on the identification of different classes of user behavior to improve recommendation systems for advertisements in OSNs. Cha et al. (2007) analyze properties of the user generated videos, such as popularity shifts, whereas parameters like video traffic, file attributes and bit-rate, are studied by Gill et al. (2007). A study on the age differences and similarities of users w.r.t. their activities in MySpace is described by Pfeil et al. (2009), who explore potential differences in social capital among older people compared to teenagers.

Several studies have looked at the comparison of the behavior of sample of users among different OSNs (Gyarmati and Trinh, 2010; Zhao et al., 2011; Dwyer et al., 2007; Shen et al., 2013; Hughes et al., 2012; Ahn et al., 2007; Fogg and Iizawa, 2008; Gao et al., 2012; Mislove et al., 2007). However, all these studies extract trends on the use of social sites and compare them through statistical parameters derived from the analysis of large sets of users. Thus, they do not observe the behavior of the same user in the two systems. Gyarmati and Trinh (2010) try to characterize user activities and usage patterns in some popular OSNs like Bebo, MySpace, Netlog, and Tagged. In Zhao et al. (2011), instead, the authors consider the differences between Twitter and traditional news media content. A comparison of Facebook and MySpace on the aspects of trust and privacy is reported by Dwyer et al. (2007). The obtained results show that members of both sites have similar levels of privacy concern. However, Facebook members report higher trust in both the social network itself and the other Facebook users, and are more confident to share identifying information. Shen et al. (2013) collect objective, privacy-preserved behavior data from user that are active in both Facebook and Gmail. The authors make a comparative analysis on user behavior in OSNs and their way of using email services. The analysis shows that a large portion of social interactions still occur through email messages, whereas participants tend to be more emotional on Facebook.

Using a general population sample of 300 users, Hughes et al. (2012) examine the personality correlates (i.e., Neuroticism, Extraversion, Openness-to-Experience, Agreeableness, Conscientiousness, Sociability and Need-for-Cognition) of social and informational use of the two OSNs Facebook and Twitter. By examining also age and gender they show that personality is related to online socializing and on how people seek and/or exchange information. Moreover, a preference for Facebook or Twitter is associated with differences in personality. Ahn et al. (2007) analyze sample networks from Cyworld, Orkut, and MySpace in terms of degree distribution, clustering coefficient, degree correlation, and average path length. Fogg and Iizawa (2008) analyze the role of persuasion in the actions that users perform in two social networking sites. The samples analyzed comprise U.S. users from Facebook and Japan users from Mixi. The authors compare the two OSNs on four persuasion goals: creating profile pages, inviting friends, responding to content by friends, and how frequently they connect to the site. Their analysis reveals the differences and similarities in how Facebook and Mixi are designed to influence users towards the achievement of these four goals.

Gao et al. (2012) compare user behavior on Sina Weibo and Twitter, by analyzing how people access microblogs and by comparing the writing style of Sina Weibo and Twitter through textual features of microposts. Moreover, based on semantics, they study and compare the topics and sentiment polarities of posts of the two systems. Finally, they investigate the temporal dynamics of the microblogging behavior such as the drift of user interests over time. Núñez-Valdéz et al. (2012) perform an analysis to find correlations between explicit and implicit feedback on recommender systems. Jin et al. (2013) study user behavior from four different perspectives, *connection* and *interaction*, *traffic activity*, *mobile social behavior*, and *malicious behaviors* in order to classify attacks (such as spam and Sybil attack) on the basis of the severity level of security threat.

Our approach differs from those presented in this section because, while all of them base their analysis on samples derived from different social networks and generalize on user behavior by observing statistical aggregative parameters, our analysis, instead, is carried out on a set of pairs of accounts (of `Twitter` and `Facebook`, respectively) such that the accounts of each pair belong to the same user. Hence, our study actually reflects the different way of being and acting of the same person in the two considered social networks. To the best of our knowledge, this paper is the first study addressing the problem from this perspective.

3. Materials and Methods

The specificity of our analysis is to consider accounts of the same person in `Twitter` and `Facebook` in order to draw conclusions on the use of the two social networks by highlighting similarities and differences. In the following, we provide the detail to allow full reproducibility of all experiments.

Data Extraction. Information necessary for our analysis has been extracted from `Twitter` and `Facebook` from January to May 2014, via three technologies:

- APIs provided by the social networks themselves;
- supplementary FOAF datasets;
- HTML parsing.

APIs are a platform available for developers which allow the access to social-networks data so as to create applications on top of them. Usually, there are different kinds of APIs each providing specific services. Among them, the most commons are the REST API, the Search API and the Streaming API. Specifically, the REST APIs allow operations such as insert, update or deletion to be performed. The Search APIs, instead, are useful to query the database and, finally, the Streaming APIs are designed for applications that need to receive real-time updates (such as, new posts or feeds). As for the second technology, it relies on the FOAF project (Brickley and Miller, 2000), which focuses on the creation of a machine-readable ontology describing friendship relationships among users. FOAF data sources allow the representation of a whole social network without the need of a centralized database. As a matter of fact, by relying on this technology, it is possible to represent the information concerning a user account, along with the corresponding contacts and activities, through an RDF graph serialized as an XML document, according to the W3C RDF/XML syntax. The last data extraction solution leverages on HMTL parsing. Processing HTML to obtain social data is the most intricate procedure. Parsing requires much time because it needs to analyze all context information from the page source code. It is a low-level way of dealing with social data. Because the code written depends on the HTML page structure, it is not stable (due to the frequent graphical changes). For this reason, this strategy needs continue maintenances.

An important issue in the extraction of our data is the need to detect whether two accounts belong to the same person. Fortunately, users can explicitly declare connections from the profile of a social network to another by means of special links, called `me` edges. From a technological points of view, there exist several ways to extract this information from the account of a user in a social network. The most common leverages on XFN (*XHTML Friends Network*) which is an HTML microformat allowing for the representation of the kind of relationship existing between two user accounts. This is obtained by empowering the set of values that the `rel` attribute of the HTML tag `<a>` (which represents a link) can assume. In our case, we focus on the value "me" (`rel='me'`) which indicates that the corresponding link represents a `me` edge. Another common way for extracting information on `me` edges relies, once again, on the use of social network APIs already mentioned above.

Data Sampling. The major problem in this task was the need of collecting data about user accounts that have a `me` edge from a social network to the other. To do this, visiting a social graph by any existing crawling technique results in biased data, thus it is not suitable. Indeed, it has been deeply studied that classical techniques for sampling a social graph, such as Breadth First Search and Random Walk, produce samples biased in the node degree distribution (Gjoka et al., 2010; Kurant et al., 2010) and newest sampling strategies, such as Metropolis-Hasting random walk and re-weighted random walk, solve the above problem of node degree distribution but still produce samples with very few *me* edges, as proved in (Buccafurri et al., 2014b).

As a consequence, we decided to perform uniform sampling, as it has been referred as the ground truth technique for obtaining unbiased social network datasets (Gjoka et al., 2010). Uniform sampling is not a trivial task in general. However, for `Facebook` and `Twitter`, this activity is facilitated by how user identifiers are defined. Indeed, both adopt 64-bit identifiers for user accounts. In particular, the URL address of the profile page of a `Facebook` (resp. `Twitter`) user is `http://www.facebook.com/YYY` (resp., `http://twitter.com/account/redirect_by_id?id=YYY`), where YYY is a 64-bit positive integer. Thus, to obtain a uniform sample, it suffices to generate numbers uniformly at random in a suitable interval and, for each number, to verify whether it corresponds to an existing account (because an account could have been deleted).

Extraction, Transformation and Loading of data. As we were interested in data about users who have accounts in both social networks, uniform sampling has been executed as follows. We started by uniformly sampling

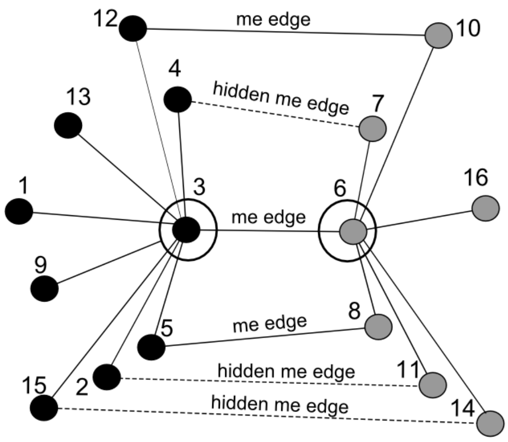
Twitter to collect a set of 875 Twitter users declaring (by a me edge) an account also in Facebook. Then, we proceeded by visiting their Facebook accounts and we found that 118 of them were not valid URLs, therefore we cleaned up our dataset by removing these nodes. For the 757 remaining Twitter bridges, we gathered information about their alternative accounts on Facebook and about their direct neighbors. The dataset obtained is composed by the following tables: user, friend and me. The first table contains the fields: screen-name, indegree, outdegree, sn_id, social network, visited, public. The first attribute screen-name is the account name chosen by the user at the moment of the registration; indegree represents the number of *followers* of the user; outdegree, instead, is the number of *followings* of the user[1]; sn_id is the original social network identification for the user; the field social network can assume two values, namely Facebook or Twitter, specifying the referring social network; finally, the attributes visited, public are two binary values indicating whether the profile has been directly visited or if it has been found as friend of a visited profile, and whether the profile has accessible information (public) or not. The table friend represents the social graph, i.e. maps the friendships of the accounts sampled, whereas table me contains information about me edges. After the collection of rough data from the previous steps, we need to preprocess data by removing duplicates and accounts with not valid URLs and by generating tables indexes and keys. To support our analysis, in a second round, we added two additional fields at the table users, namely tweet count and creation date. These new attributes have not null values only for Twitter users.

Observe that, in a previous study, Buccafurri et al. (2012) showed that, due to disparate reasons, users often do not declare *me* edges explicitly and proposed an algorithm to infer hidden me edges between two accounts. Hence, we built a further table, namely hidden me with the results of the application of the technique described by Buccafurri et al. (2012), to find further pairs of accounts associated with the same user.

Dataset Description. In order to better understand the aspects described above (e.g., me edges, hidden me edges, etc.), a portion of our dataset related to a user is sketched in Figure 1: black and gray nodes are user accounts of Twitter and Facebook, respectively, and the real name of the account is also reported[2]. Specifically, we consider a user having an account on Twitter (node 3) who declared a me edge to his Facebook account (node 6) and his social network friends (neighborhoods). In this case, also some neighbor accounts are overlapping: in particular, nodes 12 and 10, nodes 5 and 8, belong to the same user because it is explicitly declared by means of me edges; whereas, nodes 4 and 7, nodes 2 and 11, nodes 15 and 14, have been found to be of the same user by

[1]Due to the symmetric nature of the friendship relationship in Facebook, indegree and outdegree have the same value for Facebook users.

[2]Data were not anonymized at this stage to preserve full reproducibility of experiments. In case of publication, we will do this.



Id	Screen-name
1	tatahsantana
2	neltonpyter
3	Pedao_Vini
4	denioneto
5	Raafinha_mlk
6	pedro.vinicios.372
7	denio.neto
8	Raafinha.S
9	TwitPic
10	carlos.junior.50767
11	neltonpyter2
12	Carlos_Juniiior
13	instagrao
14	gabrielabritto1
15	Gabrielabritto1
16	monalisa.carla

Figure 1: A fragment of our dataset.

the algorithm proposed by Buccafurri et al. (2012), and, therefore, are labeled as `hidden me` edges (i.e., not explicitly declared).

The whole dataset is available at the address `http://www.infolab.unirc.it/cihb2014.html` (for the Reviewers: the password to open the archive is 081733849) and some statistics are reported in Table 1.

4. Results

In this section, we perform a number of experiments on the collected sample to answer the questions presented in the introduction.

4.1. Privacy setting

The first analysis concerns the choice users about the privacy level in `Facebook`. We investigate if users who have two accounts (in `Twitter` and `Facebook`) show the same behavior as other users when it comes of privacy concerns. Therefore, in this experiment the control variable is the user having two accounts, whereas the dependent variable is the privacy setting. We count how many users of the sample with two accounts choose to disclose their `Facebook` information on the social network, thus making their `Facebook` account public[3]. We obtain

[3]This analysis is limited to `Facebook` because `Twitter` accounts cannot be private.

Table 1: Some statistics of our dataset.

Seen Nodes	$\sim 4 \cdot 10^6$
Visited Nodes	304,715
`Twitter` *Nodes*	*158,755*
`Facebook` *Nodes*	*145,960*
Visited Edges	368,314
Bridges	910
`Twitter` Id Range	[1 - 359999]
Indegree Range	[0 - 52,295,363]
Outdegree Range	[0 - 2,436,264]
Bridge Indegree Range	[1 - 238,523]
Outdegree Range	[1 - 99,843]

Table 2: The logarithmic binning function used to discretize degree.

Value	Bin
$x < 10$	1
$10 \leq x < 100$	2
$100 \leq x < 1000$	3
$1000 \leq x < 10,000$	4
$x \geq 10,000$	5

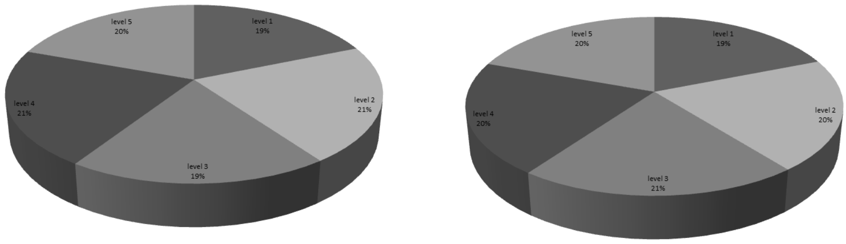
that about 87% users kept their `Facebook` account private, and this result is statistically valid with a 95% confidence level and 2.25% margin of error.

Moreover, we analyze if there are some differences, in terms of privacy setting, among users with different number of friends (i.e., degree). In this case, we consider also the degree as independent variable. To perform this analysis, we discretize degree by applying the logarithmic binning function reported in Table 2. The choice of the logarithmic binning function allows us to obtain almost equal-width bins (Milojević, 2010) due to the well-known power law distribution of node degree (Lu and Wang, 2014; Buccafurri et al., 2013).

In Figure 2, we report the distribution of the users with private account according to their discretized degree (indegree and outdegree in Figure 2.(a) and (b), respectively). We observe that there are no significant differences among the five degree intervals considered in our experiment: indeed, about one fifth of private accounts are in each of the five bins. We can conclude that the number of friends (i.e., the node degree) does not seem to affect the choice of having a private account.

4.2. Friend overlap

In this section, we study the attitude of users to have overlapping friendship relations. In particular, we analyze if users with account both in `Twitter` and `Facebook` have overlapping neighborhoods in these two social networks (i.e., how often they add the same person as friend both in `Twitter` and `Facebook`).



(a) *Indegree* (b) *Outdegree*

Figure 2: Private account distribution on the basis of their degree.

Table 3: Computing the friend overlap in `Twitter` and `Facebook`.

	M	SD
CFF_T	0.028	0.048
CFF_F	0.007	0.015
CFF_T^*	0.075	0.074
CFF_F^*	0.029	0.049

In this experiment, given a user u with account in both `Twitter` and `Facebook`, we define the metric CFF_T (*Common Twitter Friend Fraction*), which measures the fraction of the friends of u in `Twitter` who are also friends of u in `Facebook`. Analogously, we define CFF_F[4], which considers `Facebook` instead of `Twitter`. To detect if a friend is the same in the two social networks, we check for the presence of a `me` edge between these two accounts, as described in Section 3. Observe that it could occur that a friend has not explicitly declared the `me` edge and this could lead to underestimate the friend overlap. To overcome this problem, we use the approach proposed by Buccafurri et al. (2012) for discovering not declared `me` edges, which allows us to detect (with a good approximation) also these overlapping friends. We denote by CFF^* the results of the computation of friend overlapping obtained by extending the set of common accounts with the approach proposed by Buccafurri et al. (2012). Therefore, in this experiment we consider the user with two accounts and the social network as control variables; whereas *Common Friend Fraction* is the dependent variable.

The result of this analysis is reported in Table 3, in which the metrics defined above were summed and averaged and the standard deviation is also computed. This experiment shows that there is no significant overlap among the friends of the two accounts of a user in `Twitter` and `Facebook`. Indeed, the overlap is only about 7% in `Twitter` on average, and the overlap measured on `Twitter` is higher than on `Facebook`.

[4]Clearly, $CFF_T \neq CFF_F$ because the initial user set from which the fraction is computed is different.

Table 4: Average number of Twitter and Facebook friends.

	Twitter	Facebook
Indegree	1626.92	133.71
Outdegree	587.43	133.71

4.3. Friend distribution

The aim of this section is to study the relation between number of friends and membership overlap between Twitter and Facebook, by means of three experiments.

In the first experiment, we consider each user with account both in Facebook and Twitter and compare the number of friends he has in Twitter and Facebook. Therefore, the control variables of this study are the user with two accounts and the social network, whereas the dependent variables are the indegree and outdegree of users. The result of this measure is reported in Table 4, which shows the average value of indegree and outdegree of Twitter accounts versus the degree of Facebook accounts of the same users. Observe that, for Facebook, indegree and outdegree coincide because of the symmetric friendship relation. This table shows that the average degree of Twitter accounts is much higher than that of Facebook. However, because it is well-know that degree in social networks follows a power law distribution (Lu and Wang, 2014; Buccafurri et al., 2013), we need to better investigate this results.

For this purpose, in the second experiment, we compute the median value (i.e., the central value separating the higher half of degree values from the lower half) instead of the average degree, as the former is a more meaningful indicator of the trend of degree in case of power law distribution. We partition the users of our sample into the following four sets, and for each of them we compute the median value of degree:

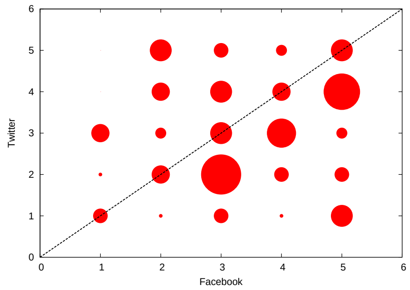
1. declaredT, composed of the Twitter users who declared to have an account also in Facebook;
2. otherT, the remaining Twitter users;
3. declaredF, composed of the Facebook users who declared to have an account also in Twitter;
4. otherF, the remaining Facebook users;

The results of this experiment are shown in Table 5. Combining these results with those of the previous experiment, we find that, while most of the users of Twitter have a degree lower than the users of Facebook (see Table 5), Twitter *power users* (i.e., the users with a very large number of friends) have a degree much higher than *power users* of Facebook (Table 4).

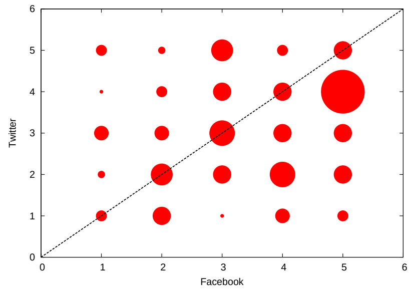
In the last experiment, we consider again a user with account both in Twitter and Facebook, and we study a possible relation between the number of friends he has in each social network. For this purpose, we discretize the indegree and outdegree in 5 equal-width bins by applying the logarithmic

Table 5: The median of indegree and outdegree for the four sets of users.

	indegree	outdegree
`declaredT`	61	114
`otherT`	53	108
`declaredF`	726	726
`otherF`	679	679



(a) *Indegree* `Twitter` *vs Degree* `Facebook`.



(b) *Outdegree* `Twitter` *vs Degree* `Facebook`.

Figure 3: The scatter plot showing the relation between number of friends in `Twitter` and `Facebook`.

binning function reported in Table 2 and we build two dispersion matrices having the indegree (resp., outdegree) level of `Twitter` as Y-axis and the degree of `Facebook` as a X-axis. This way, we can observe if a relation between degrees of the same user in the two social networks exists. Figure 3 shows the graphical representation of the two dispersion matrices: the more the points are next to the bisecting line, the higher the relation between the degrees. Moreover, the size of the points represents the number of occurrences of that degree combination.

From the analysis of this figure, we observe that the bigger points are in the area under the bisector, meaning that there is a slight anti-correlation between `Facebook` and `Twitter` degrees (i.e., the higher the degree of a user in `Facebook`, the lower the corresponding degree in `Twitter`).

4.4. User Activity

Through the experiments described in this section, we study the relation about membership overlap and user activity. Specifically, the purpose of this analysis is to investigate the behavior of three different typologies of users: (1) `declared`, which are users who have an account both in `Facebook` and `Twitter` and explicitly declared this, (2) `hidden`, which are users who have an account both in `Facebook` and `Twitter` but did not declare this[5], (3) `other`, which are

[5] As done in Section 4.1, such users are detected by using the approach defined by Buccafurri et al. (2012).

Table 6: The frequency distribution of the normalized activity coefficient for the different types of user.

(a) `Declared` users.

NAC bin	%
3	32.4
1	27.5
2	24.9
4	13.8
5	1.3

(b) `Hidden` users.

NAC bin	%
1	91.3
3	3.8
4	3.6
2	0.7
5	0.3

(c) `Other` users.

NAC bin	%
4	46.1
3	34.0
5	9.0
2	7.2
1	3.4

the remaining users.

For this analysis, we define the normalized activity coefficient NAC of a `Twitter` account as t_c/y_a, where t_c is the number of tweet posted and y_a is the number of years since the account has been created. As done for degree in Section 4.1, we discretize NAC by applying the logarithmic binning function reported in Table 2 and we compute its value for each typology of users. In this study, the user typology is our control variable, whereas the activity coefficient is the dependent variable. The obtained results are reported in Table 6, in which the first row represents the mode (i.e., the bin which most of the users fall in) for the specific type of user. From these results, we find that users who have more accounts are less active than the others who have account only in Twitter. By looking at Figure 4, which shows a different view of the results, we conclude that, among users with more accounts, those who have a not declared Facebook account are very inactive.

5. Discussion

The goal of our paper is to analyze the behavior of users belonging to more social networks. The chosen setting refers to the two most popular social networks, which are `Facebook` and `Twitter`. In this section, we highlight and discuss our major findings and issues.

The first analysis we have carried out concerns privacy awareness. We have found that 87% users with account in `Facebook` and `Twitter` chose to keep their `Facebook` information private (i.e., they have a private accounts and their information are accessible only to their friends). Moreover, we observed that this percentage is not significantly affected by the number of friends they have. As precedent studies have shown that only 52% `Facebook` users have private accounts (Dey et al., 2012), we conclude that the users declaring to have an account both on `Facebook` and `Twitter` pay more attention to privacy issues. These users are able to declare a secondary account (i.e., by a `me` edge), know technological aspects of the social network, and are aware of privacy setting. As a consequence, they may be more accurate in the definition of the privacy policy for their accounts. The first important result of our study is that privacy

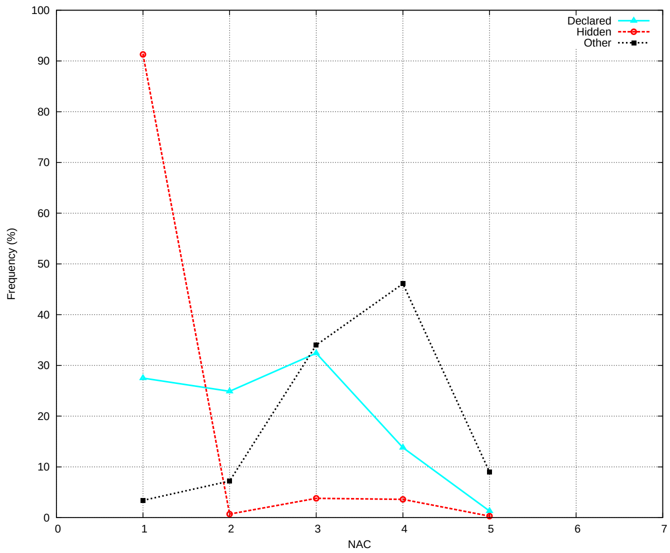


Figure 4: The frequency distribution of the normalized activity coefficient for the different types of user.

awareness has a positive impact on privacy value. This result confirms a previous study done by Cranor et al. (2008), in the context of e-commerce, financial and healthcare websites, which showed that the more the users are informed about privacy, the less they self-disclose. Moreover, Stutzman et al. (2011) found that the users who have customized privacy settings are less likely to disclose information.

The second result we found is that there is no significant overlap among the friends of a user in `Twitter` and `Facebook`. This behavior can be explained by considering that, in general, users create multiple accounts in social networks for different purposes (e.g., sport profiles, music profile, job profiles, etc.). Thus, each account is associated with a specific interest or part of the user life and the friends they add comprise people related to the specific context which the profile refers to. A recent study by Hughes et al. (2012) showed that personality differences between `Facebook` and `Twitter` users exist, such that more sociable individuals gravitate towards `Facebook`, while less sociable ones gravitate towards `Twitter`, thus making friendship overlap less likely. Our result confirms the study done by Panek et al. (2013) according to which `Facebook` users tend to add as friend people they know in real life in order to transform latent to weak ties (Ellison et al., 2007), whereas `Twitter` use is driven primarily by interest for entertainment news, celebrity news, and sports news (Hargittai and Litt, 2011). Moreover, our findings confirm also the hypothesis that the notion of online friend can comprise different kinds of friendship, as suggested by Tong et al. (2008).

The third analysis concerned the number of friends of `Twitter` and `Facebook` users. The result obtained about the median and average number of friends in `Facebook` is consistent with that reported in the study of Bohn et al. (2014) and confirms that the number of (stable) friends is bounded (around 150) due to limitations of the human brain (Dunbar, 2012). We found that, while most of the users of `Twitter` have a degree lower than the users of `Facebook`, `Twitter` *power users* have a degree much higher than *power users* of `Facebook` (Table 4). This can be explained according to the theory that follow relationships in `Twitter` are typically towards important and famous people who act as *power users* (Hargittai and Litt, 2011), so that there are few users but with a very large degree. In contrast, in `Facebook`, friends are often personally known (Ellison et al., 2007) so that their number is limited. Finally, our result about anti-correlation between degree in `Twitter` and `Facebook` allows us to conclude that a user joining both `Twitter` and `Facebook` does not equally subdivide his activity between the two social networks, but has a preference for one. From the analysis of our sample, we observed that such users prefer `Facebook` as main platform.

The findings of the last experiment is that users who have more accounts are less active than the others who have account only in `Twitter`. This can be explained by considering that the latter users may focus their attention only in one social network, thus directing all their posting activities on it. Vice versa, the users posting contents in at least two social networks (i.e., `Twitter` and `Facebook`), concentrate the total amount of posts in one of them. Moreover, we found that users who do not declare to have more accounts are the least active and appear "lazy". The importance of this result is related to several aspects of user behavior: it has been found that people who are active on social networks are more likely to feel connected (Ellison et al., 2007; Valenzuela et al., 2009; Steinfield et al., 2008; Chen, 2011), that they are high in ICT innovativeness (Zhong et al., 2011) and that social activity may enhance social presence and increase social influence (Cheung et al., 2011).

6. Conclusion and Limitations

Social network analysis has assumed an extraordinary importance since the birth of online social networks, because they target and record social relationships in the most complete and detailed way among all digital services. As a consequence, they are a huge source of information about people, whose analysis may result in strategic and useful knowledge. The multiple-social-network perspective is a promising approach compliant with the evolution of social web. In this paper, we have addressed the problem of comparing the behavior of a user in `Twitter` and `Facebook`, believing that the multiplicity of social networks is an important aspect to take into strict consideration when studying the complex phenomenon of OSNs. The novelty of this paper is to have approached the comparison from a truly multiple-social-network perspective, according to which users with profiles in both OSNs have been considered as object of study, to have meaningful and well-founded results. This study has some limitations.

First, it considers only `Facebook` and `Twitter` among the numerous online social networks, so that our results can not be generalized to other or all social networks. However, we chose the two most popular ones, thus ensuring the validity of results for a very large part of social network users. The next limitation concerns the extraction of the sample from the Web, because only public information can be retrieved, and the limited size of the sample. Another limitation is that, like many other observational studies, we cannot draw causal conclusions. To overcome these limitations, future work could extend the sample size both considering other OSNs and using other extraction techniques, such as surveys, yet taking into account that this method results in a strong sampling bias and makes it difficult to acquire large samples. Moreover, because some results could be explained by other latent variables, such as age, location, sex, these should be incorporate in the future analysis.

References

Ahn, Y.-Y., Han, S., Kwak, H., Moon, S., Jeong, H., 2007. Analysis of topological characteristics of huge online social networking services. In: Proceedings of the 16th international conference on World Wide Web. ACM, pp. 835–844.

Barash, V., Ducheneaut, N., Isaacs, E., Bellotti, V., 2010. Faceplant: Impression (mis) management in facebook status updates. In: ICWSM.

Benevenuto, F., Rodrigues, T., Cha, M., Almeida, V., 2009. Characterizing user behavior in online social networks. In: Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference. ACM, pp. 49–62.

Bergman, S. M., Fearrington, M. E., Davenport, S. W., Bergman, J. Z., 2011. Millennials, narcissism, and social networking: What narcissists do on social networking sites and why. Personality and Individual Differences 50 (5), 706–711.

Bohn, A., Buchta, C., Hornik, K., Mair, P., 2014. Making friends and communicating on facebook: Implications for the access to social capital. Social Networks 37, 29–41.

Brickley, D., Miller, L., 2000. FOAF Vocabulary Specification 0.91. Tech. rep., Tech. rep. ILRT Bristol. Available at `http://xmlns.com/foaf/spec/20071002.html`.

Buccafurri, F., Foti, V., Lax, G., Nocera, A., Ursino, D., 2013. Bridge Analysis in a Social Internetworking Scenario. Information Sciences 224, 1–18, elsevier.

Buccafurri, F., Lax, G., Nicolazzo, S., Nocera, A., Ursino, D., 2014a. Driving Global Team Formation in Social Networks to Obtain Diversity. In: Proc. of the International Conference on Web Engineering (ICWE 2014). Toulouse, France.

Buccafurri, F., Lax, G., Nocera, A., 2015. A New Form of Assortativity in Online Social Networks. International Journal of Human-Computer Studies-Forthcoming.

Buccafurri, F., Lax, G., Nocera, A., Ursino, D., 2012. Discovering Links among Social Networks. In: Proc. of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2012). Lecture Notes in Computer Science. Springer, Bristol, United Kingdom, pp. 467–482.

Buccafurri, F., Lax, G., Nocera, A., Ursino, D., 2014b. Moving from Social Networks to Social Internetworking Scenarios: the Crawling Perspective. Information Sciences 256, 126–137.

Burke, M., Kraut, R., Marlow, C., 2011. Social capital on facebook: Differentiating uses and users. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, pp. 571–580.

Burke, M., Marlow, C., Lento, T., 2009. Feed me: motivating newcomer contribution in social network sites. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, pp. 945–954.

Carpenter, C. J., 2012. Narcissism on facebook: Self-promotional and anti-social behavior. Personality and individual differences 52 (4), 482–486.

Cha, M., Kwak, H., Rodriguez, P., Ahn, Y.-Y., Moon, S., 2007. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement. ACM, pp. 1–14.

Chen, G. M., 2011. Tweet this: A uses and gratifications perspective on how active twitter use gratifies a need to connect with others. Computers in Human Behavior 27 (2), 755–762.

Cheng, X., Dale, C., Liu, J., 2008. Statistics and social network of youtube videos. In: Quality of Service, 2008. IWQoS 2008. 16th International Workshop on. IEEE, pp. 229–238.

Cheung, C. M., Chiu, P.-Y., Lee, M. K., 2011. Online social networks: Why do students use facebook? Computers in Human Behavior 27 (4), 1337–1343.

Cranor, L. F., Egelman, S., Sheng, S., McDonald, A. M., Chowdhury, A., 2008. P3p deployment on websites. Electronic Commerce Research and Applications 7 (3), 274–293.

Dey, R., Jelveh, Z., Ross, K., 2012. Facebook users have become much more private: A large-scale study. In: Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on. IEEE, pp. 346–352.

Dunbar, R., 2012. Social cognition on the internet: testing constraints on social network size. Philosophical Transactions of the Royal Society B: Biological Sciences 367 (1599), 2192–2201.

Dwyer, C., Hiltz, S., Passerini, K., 2007. Trust and privacy concern within social networking sites: A comparison of facebook and myspace. AMCIS 2007 Proceedings, 339.

Ellison, N. B., Steinfield, C., Lampe, C., 2007. The benefits of facebook "friends:" social capital and college students' use of online social network sites. Journal of Computer-Mediated Communication 12 (4), 1143–1168.

Fogg, B., Iizawa, D., 2008. Online persuasion in facebook and mixi: A cross-cultural comparison. In: Persuasive technology. Springer, pp. 35–46.

Gao, Q., Abel, F., Houben, G.-J., Yu, Y., 2012. A comparative study of users microblogging behavior on sina weibo and twitter. In: User modeling, adaptation, and personalization. Springer, pp. 88–101.

Gill, P., Arlitt, M., Li, Z., Mahanti, A., 2007. Youtube traffic characterization: a view from the edge. In: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement. ACM, pp. 15–28.

Gjoka, M., Kurant, M., Butts, C. T., Markopoulou, A., 2010. Walking in Facebook: A case study of unbiased sampling of OSNs. In: INFOCOM, 2010 Proceedings IEEE. IEEE, pp. 1–9.

Gyarmati, L., Trinh, T. A., 2010. Measuring user behavior in online social networks. Network, IEEE 24 (5), 26–31.

Hargittai, E., Litt, E., 2011. The tweet smell of celebrity success: Explaining variation in twitter adoption among a diverse group of young adults. New Media & Society 13 (5), 824–842.

Hassan, N. R., 2009. Using social network analysis to measure it-enabled business process performance. Information Systems Management 26 (1), 61–76.

Hughes, D. J., Rowe, M., Batey, M., Lee, A., 2012. A tale of two sites: Twitter vs. facebook and the personality predictors of social media usage. Computers in Human Behavior 28 (2), 561–569.

Java, A., Song, X., Finin, T., Tseng, B., 2007. Why we twitter: understanding microblogging usage and communities. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis. ACM, pp. 56–65.

Jin, L., Chen, Y., Wang, T., Hui, P., Vasilakos, A. V., 2013. Understanding user behavior in online social networks: A survey. IEEE Communications Magazine 51 (9), 144–150.

Johnston, K., Tanner, M., Lalla, N., Kawalski, D., 2013. Social capital: the benefit of facebook friends. Behaviour & Information Technology 32 (1), 24–36.

Krishnamurthy, B., 2009. A measure of online social networks. In: Communication Systems and Networks and Workshops, 2009. COMSNETS 2009. First International. IEEE, pp. 1–10.

Kurant, M., Markopoulou, A., Thiran, P., 2010. On the bias of BFS (breadth first search). In: 22nd International Teletraffic Congress, ITC 2010, Amsterdam, The Netherlands, September 7-9, 2010. pp. 1–8.

Leskovec, J., Adamic, L. A., Huberman, B. A., 2007. The dynamics of viral marketing. ACM Transactions on the Web (TWEB) 1 (1), 5.

Lu, J., Wang, H., 2014. Variance reduction in large graph sampling. Information Processing & Management 50 (3), 476–491.

Maia, M., Almeida, J., Almeida, V., 2008. Identifying user behavior in online social networks. In: Proceedings of the 1st workshop on Social network systems. ACM, pp. 1–6.

Milojević, S., 2010. Power law distributions in information science: Making the case for logarithmic binning. Journal of the American Society for Information Science and Technology 61 (12), 2417–2425.

Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., Bhattacharjee, B., 2007. Measurement and analysis of online social networks. In: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement. ACM, pp. 29–42.

Nguyen, B., Mallet, V., Woods, N., Cheng, J., Chokshi, S., Ramanarayanan, G., Streich, M., Jun. 25 2013. Content sharing using notification within a social networking environment. US Patent 8,473,550.

Núñez-Valdéz, E. R., Lovelle, J. M. C., Martínez, O. S., García-Díaz, V., de Pablos, P. O., Marín, C. E. M., 2012. Implicit feedback techniques on recommender systems applied to electronic books. Computers in Human Behavior 28 (4), 1186–1193.

Panek, E. T., Nardis, Y., Konrath, S., 2013. Mirror or megaphone?: How relationships between narcissism and social networking site use differ on facebook and twitter. Computers in Human Behavior 29 (5), 2004–2012.

Pempek, T. A., Yermolayeva, Y. A., Calvert, S. L., 2009. College students' social networking experiences on facebook. Journal of Applied Developmental Psychology 30 (3), 227–238.

Pfeil, U., Arjan, R., Zaphiris, P., 2009. Age differences in online social networking–a study of user profiles and the social capital divide among teenagers and older users in myspace. Computers in Human Behavior 25 (3), 643–654.

Rodriguez, P., April 2009. Web Infrastructure for the 21st Century. In: 18th International World Wide Web Conference.
URL http://www2009.eprints.org/213/

Rosen, L. D., Whaling, K., Rab, S., Carrier, L. M., Cheever, N. A., 2013. Is facebook creating idisorders? the link between clinical symptoms of psychiatric disorders and technology use, attitudes and anxiety. Computers in Human Behavior 29 (3), 1243–1254.

Ross, C., Orr, E. S., Sisic, M., Arseneault, J. M., Simmering, M. G., Orr, R. R., 2009. Personality and motivations associated with facebook use. Computers in Human Behavior 25 (2), 578–586.

Shen, J., Brdiczka, O., Ruan, Y., 2013. A comparison study of user behavior on facebook and gmail. Computers in Human Behavior 29 (6), 2650–2655.

Staddon, J., Huffaker, D., Brown, L., Sedley, A., 2012. Are privacy concerns a turn-off?: engagement and privacy in social networks. In: Proceedings of the eighth symposium on usable privacy and security. ACM, p. 10.

Steinfield, C., Ellison, N. B., Lampe, C., 2008. Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. Journal of Applied Developmental Psychology 29 (6), 434–445.

Stutzman, F., Capra, R., Thompson, J., 2011. Factors mediating disclosure in social network sites. Computers in Human Behavior 27 (1), 590–598.

Teevan, J., Ramage, D., Morris, M. R., 2011. # twittersearch: a comparison of microblog search and web search. In: Proceedings of the fourth ACM international conference on Web search and data mining. ACM, pp. 35–44.

Tong, S. T., Van Der Heide, B., Langwell, L., Walther, J. B., 2008. Too much of a good thing? the relationship between number of friends and interpersonal impressions on facebook. Journal of Computer-Mediated Communication 13 (3), 531–549.

Valenzuela, S., Park, N., Kee, K. F., 2009. Is there social capital in a social network site?: Facebook use and college students' life satisfaction, trust, and participation1. Journal of Computer-Mediated Communication 14 (4), 875–901.

Watts, D. J., Peretti, J., Frumin, M., 2007. Viral marketing for the real world. Harvard Business School Pub.

Williamson, D. A., 2007. Social network marketing: Ad spending and usage. Social Network Marketing, Report by Debra Aho Williamson.

Wilson, C., Boe, B., Sala, A., Puttaswamy, K. P., Zhao, B. Y., 2009. User interactions in social networks and their implications. In: Proceedings of the 4th ACM European conference on Computer systems. Acm, pp. 205–218.

Zhang, H., Nguyen, D. T., Zhang, H., Thai, M. T., 2015. Least cost influence maximization across multiple social networks.

Zhang, J. X., Zhang, H., Ordóñez de Pablos, P., Sun, Y., 2014a. Challenges and foresights of global virtual worlds markets. Journal of Global Information Technology Management 17 (2), 69–73.

Zhang, X., de Pablos, P. O., Wang, X., Wang, W., Sun, Y., She, J., 2014b. Understanding the users' continuous adoption of 3d social virtual world in china: A comparative case study. Computers in Human Behavior 35, 578–585.

Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., Li, X., 2011. Comparing twitter and traditional media using topic models. In: Advances in Information Retrieval. Springer, pp. 338–349.

Zhong, B., Hardin, M., Sun, T., 2011. Less effortful thinking leads to more social networking? the associations between the use of social network sites and personality traits. Computers in Human Behavior 27 (3), 1265–1271.

Zlatolas, L. N., Welzer, T., Heričko, M., Hölbl, M., 2015. Privacy antecedents for sns self-disclosure: The case of facebook. Computers in Human Behavior 45, 158–167.