ARTICLE IN PRESS
JID: COMCOM [m5G;October 14, 2015;12:1]

Computer Communications 000 (2015) 1–8

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

The impact of user's availability on On-line Ego Networks: a Facebook analysis

Andrea De Salve [a], Marco Dondio [a], Barbara Guidi [a,b], Laura Ricci [a,*]

[a] Department of Computer Science, University of Pisa, Largo Bruno Pontecorvo, Pisa, Italy
[b] Institute for Informatics and Telematics, IIT, CNR, Via Moruzzi, Pisa, Italy

A R T I C L E I N F O

Article history:
Available online xxx

Keywords:
Online Social Networks
Temporal homophily
Dunbar's circles

A B S T R A C T

Online Social Networks (OSNs) are the most popular applications in todays Internet and they have changed the way people interact with each other. Understanding the structural properties of OSNs and, in particular, how users behave when they connect to OSNs is crucial for designing user-centered systems. Results about OSNs demonstrated that the relationships that an individual (*ego*) maintains with other people (*alters*) can be organized into a set of circles (named *Dunbar's circles*) according to the *ego network model*. The study of the impact of ego networks structure on the availability patterns of users is seriously limited by the lack of information about users availability patterns. In this work we contribute to fill this gap by analysing availability information of a sample of Facebook users. The data reveal a number of strong temporal dependencies (or *temporal homophily*) which provide insights into the availability pattern that characterize an ego network.

1. Introduction

Online Social Networks (OSNs) have become widely used in the last few years. Because of their huge impact on people's lives, OSNs became a quite interesting research field and resource for many scientists: sociologists are investigating the dynamics occurring inside OSNs and trying to compare them with real life behaviors [1–3], computer scientists and engineers are focusing on the technical side by studying their properties [4,5], improving these architectures [6–8] and optimizing their usage [9].

Users of the OSNs generate a high-volume of valuable information which are controlled and stored by the centralized providers. Although users are forced to trust the provider of the services, recent events have shown that in addition to malicious users (internal or external to the OSN), also the centralized service provider [10] introduce new security and privacy risks. However, recently, users are realizing the privacy risks deriving from giving up their data to a centralized service. For these reasons in the last years, scientists are studying new infrastructures to offer OSN services by exploiting distributed (for instance P2P) patterns. Diaspora[1], with about 400,000[2] users, is one of the most successful OSN services implemented in a decentralized way. These systems are referred as Distributed Online Social Networks (DOSNs) [11] and their development has to face a large number of issues. The main one is the data availability and persistence problem because of the lack of a central server storing users' data and because of the intrinsic dynamism of these networks. When users disconnect from the OSN, the knowledge of their friendship relationships and of the kind of interactions with their friends may be exploited to store data on trusted contacts so avoiding data encryption which would otherwise be required if data is stored on any node of the network (needed for instance in DTN or CDN-based approaches). A further challenge is to consider the availability patterns of the nodes in the ego network to define data allocation, with the goal of avoiding continual data transfers which affect the overall performance of the social service. The study of the temporal behavior of users in OSNs, in particular the study of the relation between online sessions of egos and those of their alters is therefore of primary importance to help the decentralization of social services by characterizing the typical OSNs' usage and understanding how users interact with these platforms.

In the last few years, the study of ego networks in virtual environments has received more and more attention [12]. Ego networks are social networks made up of an individual (*ego*) and all the social relations it has with other people (*alters*). Even if available OSNs datasets has led to a better understanding of OSNs at global structural level [13–15], as well as at the level of local structural properties [1,2,15,16], some characteristics of ego networks still remain unexplored, particularly their temporal characteristics because of the lack of real temporal data. Temporal properties of OSNs' users are important for different purposes. The studies of user behaviors allow the performance

* Corresponding author. Tel.: +39 0502212768.
E-mail address: laura.ricci@unipi.it, ricci@di.unipi.it (L. Ricci).

1 https://diasporafoundation.org/
2 https://diasp.eu/stats

http://dx.doi.org/10.1016/j.comcom.2015.09.001
0140-3664/© 2015 Published by Elsevier B.V.

of existing systems to be evaluated and lead to a better site design of the OSN services. Moreover, understanding ego network temporal properties in DOSNs is crucial for social studies as well as for the analysis of the dynamic processes that occur on them (such as protocol for the information diffusion [5]). Finally, understanding the temporal correlation between users sessions is valuable in designing the next-generation of OSN infrastructures, such as Distributed Online Social Networks [11] and Content Distribution Systems [6–8]. While few recent studies examined the availability patterns of OSNs' users by focusing either on the time that the user spends online [17] or on the interaction frequency [4,18], they do not provide a global picture of the relation between the ego network structure of users and the availability patterns of the alters belonging to their ego-network. The aim of this paper is to investigate the presence of temporal dependency in the ego network structure of OSNs by analysing a large set of social information about Facebook users, collected by a novel Facebook application we have developed[3]. Temporal patterns can be exploited to manage important problems in a distributed scenario, such as data availability [7] and information diffusion [8]. Using this sample of Facebook users, we find a strong relation between the ego network structure of an ego and the availability patterns of the alters in the ego network. The main result of the paper is the identification of a relation between the similarity (or *temporal homophily*) between the availability patterns of the egos and their alters, which increases when considering alters belonging to inner Dunbar circles. The remainder of this paper is organized as follows. Section 2 describes the basic concepts used and the related work Section 3 describes the data set and the methodology used to collect them. Section 4 presents the results and the methodologies used to analyze: the structural properties of the network (Section 4.1) and the interactions of the users and their Dunbar's ego network structure (Section 4.2). The Section 5 investigates the main result, that is the analysis of the temporal features of our data set. Finally, Section 6 draws the main conclusions.

2. Related work

The research community has dedicated a fair amount of work to study OSNs in the last years. This section discusses the related work mainly in the following two aspects: *(i)* existing work analysing the structural properties of ego network structure in OSNs and *(ii)* existing studies reporting the analyses of the availability pattern in OSNs.

Structural properties: Dunbar's Circles. Several studies investigate the graph structures of OSNs, such as degree distribution, clustering and ego network structure. By using the crawled data gathered from popular OSN sites, several properties have been discovered [9]. Specifically, it has been found that OSNs manifest small-world and scale-free properties. Several researchers have recently focused on the study of the structural properties of the *ego networks* in Facebook and Twitter [1,2,19], and they found that structures similar to those of offline social networks can also be identified in OSNs. As shown by Dunbar [20], there are constraints limiting the number of active relationships users can have in their ego networks. This limit is about 150 and is called the *Dunbar's number*. The active relationships of an ego can be characterized by the strength of social tie (*tie strength*): the amount of *emotional closeness* between ego and their alters [12].

As discovered by Dunbar [20], ego networks are organized into a hierarchical structure formed by four concentric subgroups of alters arranged in a concentric sequence of circles (called *Dunbar's circles*), with an increasing level of intimacy. Inner circles, called *support clique* are characterized by increasing level of tie strength while outer circles, called *sympathy group*, *affinity group* and *active network* are characterized by a low level of tie strength. After Dunbar, several studies have evaluated the presence of Dunbar's circles in different online social networks contexts and showed that it is possible to identify their average size as 5 (*support clique*), 15 (*sympathy group*), 50 (*affinity group*) and 150 (*active network*) with a scaling factor approximately equal to three [1,2,19,21].

Structural properties: tie strength. The definition of the tie strength is currently an open problem and several alternatives have been presented in the literature. Authors in [3] deduce tie strength by using a Facebook data set and explicit evaluation of tie strength done by users. Some studies such as [1,2,16] use frequency of the online interactions to estimate tie strength, since the two concepts appear to be tightly correlated. Finally, authors in [22] showed that, while frequency of interactions is a necessary component for predicting the strength of a tie, additional information related to social interactions is helpful to achieve more accurate predictions (such as number of likes, posts, comments, tags on the same picture, etc.). Finally, Sala et al. [23] exploit an indirected weighted graph to describe the social relations, because they observe that most interactions on Facebook are reciprocated so that the notion of tie strength is symmetric.

Analyses of OSN user behaviors. The works in [4,5,14,17,18] study the user behaviors by focusing on availability pattern in OSNs. Golder et al. [18] show the existence of a periodical time pattern which is influenced daily by the day/night cycle with the presence of some hourly peaks, and weekly peaks due to the weekend different habits with respect to the typical usage. Kermarrec et al. [5] use traces relative to MySpace OSN to study the correlation between availability traces of each pair of users. They observe that users are more present when their top friends are online rather than when their random friends are online. Authors in [17] analyse the workloads of the users who accessed four popular social networks. They study how frequently people connect to social networks and for how long, as well as the types and sequences of activities that users conduct on these sites, by using detailed click-stream data. Gyarmati and Trinh [14] analyse the availability pattern of users in OSNs (namely Bebo, MySpace, Netlog and Tagget). They observe that the session times of users as well as the number of sessions follow power law distributions and recently joined users may lose interest in OSNs. Jin et al. [4] conduct a comprehensive review about the availability pattern of user behavior in OSNs from several perspectives, including connectivity and interaction among users. Finally, several recent works [6,24] analyse the availability of the user to provide efficient storage strategies and data persistence in DOSN . Authors in [24] evaluate several storage and replication strategies in a Friend-to-Friend (F2F) storage system and they found that availability correlation offers a good trade-off between data availability and data redundancy. Authors in [6] exploit probabilistic models to reproduce the session characteristics of OSN users and assess their replication strategies.

Although these findings highlight some important properties of OSNs, a clear understanding of the impact of users availability on Dunbar ego networks is still missing. Moreover, to the best of our knowledge, none of the previous studies provides a detailed analysis of the availability pattern which characterize the Dunbar's circles.

3. Dataset description

The lack of data concerning online presence of users in OSNs is currently the main limitation for defining temporal pattern analysis. Furthermore, only a very small number of studies are based on complete datasets provided by the OSN operators while others have collected a complete view of specific parts of OSNs. As a matter of fact, a complete dataset is typically unavailable to researchers, as most OSNs are unwilling to share their company's data even in an anonymized form, primarily due to privacy concerns. For all these reasons, it is

[3] Available at: http://socialcircles.eu/

Please cite this article as: A. De Salve et al., The impact of user's availability on on-line ego networks: A Facebook analysis, Computer Communications (2015), http://dx.doi.org/10.1016/j.comcom.2015.09.001

common to work with small but representative samples of an OSN. At the best of our knowledge, no existing up-to-date dataset is able to provide complete information, such as information regarding the social graph, interactions among users and temporal information (on-line sessions) for a real OSN.

For these reasons we have implemented a Facebook application, called *SocialCircles!*[4], which exploits the former Facebook API (applications exploiting this API will be supported till 1st May 2015). The application is able to retrieve the following sets of information from registered users:

Topology and profile information We are able to obtain friends of registered users and the friendship relations existing between them. Moreover we download profile information of registered users and their friends, such as *complete name*, *birthday*, *sex*, *location*, *works*, *schools*, *user devices*, *movies*, *music*, *book*, *interest* and *language*.

Interaction information We have collected information about interactions between users registered to the application and their friends, such as *posts*, *comments*, *likes*, *tags* and *photo*. Due to technical reasons (time needed to fetch all data and storage capacity), we restrict the interactions information retrieved up to 6 months prior to user application registration.

Online presence data By requesting the online presence permission, we are able to monitor the *chat presence status* and obtain information about the time spent online by registered users and their friends. The chat status can assume a limited set of value: 0 if user is *offine*, 1 if user is in *active state* and 2 if user is *idle* (i.e. the user is online but they have not performed actions for more than 10 min).

The dataset obtained from the *SocialCircles!* application contains 337 complete ego networks related to the registered users, for a total of 144.481 users (ego and their alters). The resulting Facebook population has the advantage of representing a very heterogeneous population: 213 males and 115 females, with age range of 15–79 with different education, background and geographically location. We check if the sample of 337 registered users is representative of the entire population by comparing the empirical distributions of the 337 users against the distribution of the rest of the users. We use the two-sample Kolmogorov–Smirnov to compare the average session length ($p\text{-}value = 0.9992$, $D\text{-}value = 0.0011$) and the number of sessions as variables ($p\text{-}value = 1$, $D\text{-}value = 0.0002$). Given the large *p-values* and small *D-values* from the Kolmogorov–Smirnov test, we can say with high confidence that both the variables' values are drawn from the same probability distributions. We thus expect less biased results and more variability in the ego networks structure, which should better reflect real OSN utilization. Using the SocialCircles application, we aim to investigate a Facebook data set with the goals of: *(i)* study and validate important properties of OSNs, *(ii)* analyse the structure of the ego networks, *(iii)* study the availability pattern which characterizes the Dunbar ego networks in Facebook and *(iv)* analyse the extent to which the availability of an ego depends on the presence of the alters in each Dunbar's circle.

4. Analysis of the data set features

In this section we present a complete analysis of the sample data in order to assess the properties of OSNs and to compare them with the characteristics found in past research. Afterwards, we investigate the relationship existing between ego network structures and the availability of a user in the system.

4.1. Descriptive characteristics of the dataset

Initially, we analyzed the collected data from a pure topological perspective, as shown in Table 1. This analysis has been performed on the whole set of 337 ego networks. The majority (80%) of ego networks have less than 600 friends, whereas only 7% of nodes exceeds 1000 friends. Furthermore, we can notice that only 20% of ego have less than 250 friends. The distribution of values is right (positively) skewed and for this reason the median value is a better representative of the central tendency of the distribution than the mean value: we discovered that the median Facebook network has about 390 friends totally. The distribution of nodes is heavily right skewed: the very high value of standard deviation (SD) confirms that the mean value is practically useless. More than 75% of the networks have less than 10.000 ties, and the typical network exposes around 4500 connections (median value). The high value for the SD and the wide range of values suggests strong heterogeneity in our analyzed sample. A high average clustering coefficient equal to 0.636 indicates the presence of a tightly connected graph structure. We notice that the analysed sample exposes a comparable but slightly higher average clustering value with respect to past analyses (such as 0.6055 [5]). The local degree of a node is a centrality index that measures the number of mutual friends that an alter shares with each ego. It shows that ego and alters share on average 4% of friends.

To achieve a deeper understanding of the collected ego networks, we performed a clustering analysis to discover topology-based communities. We applied the Louvain community detection method [13] that aims to maximize the modularity value assigned to each node: a measure of how much a network is composed by several communities of nodes. Since ego networks contain topological information regarding ego and its alters, and no information about whole network structure and connections, we computed the centrality measure called *Ego Betweenness Centrality* (EBC). It has been defined by Everett and Borgatti in 2005 [25] as the fraction of the shortest paths that cross the ego in the ego network. The average EBC value is quite high, and seems to confirm that these ego network are composed by several different communities.

4.2. Analysis of the users interactions and ego network properties

We focused our attention on the analysis of the interactions occurring between users in our dataset. Due to the restricted access policies, we are not able to retrieve private mailbox between users but only the incoming/outgoing activity between the ego and each of its alter. We remark that regarding the outgoing activity, it was possible to retrieve only a limited subset of it: due to Facebook privacy restrictions, we could access *likes* but not *posts* nor *comments* written on friends wall, content which would need friends explicit access authorization to our application. Fig. 1(a) shows that *likes*, *photos* and *comments* contribute to over 50% of total interactions. The second most important interaction type is *comment*, which accounts for 22.7% of the overall interactions.

It has been shown that in OSN, each ego is in direct communication with less people compared to people who contact him [22]. Our analysis in fact confirms this trend, showing that the average of the incoming active network is around 29% of the alters (vs the 26% of the outgoing), although the difference is smaller compared to the work above mentioned. Furthermore, we discovered that 18% of ties in the ego network are symmetric, representing ego reciprocating the interaction with alters.

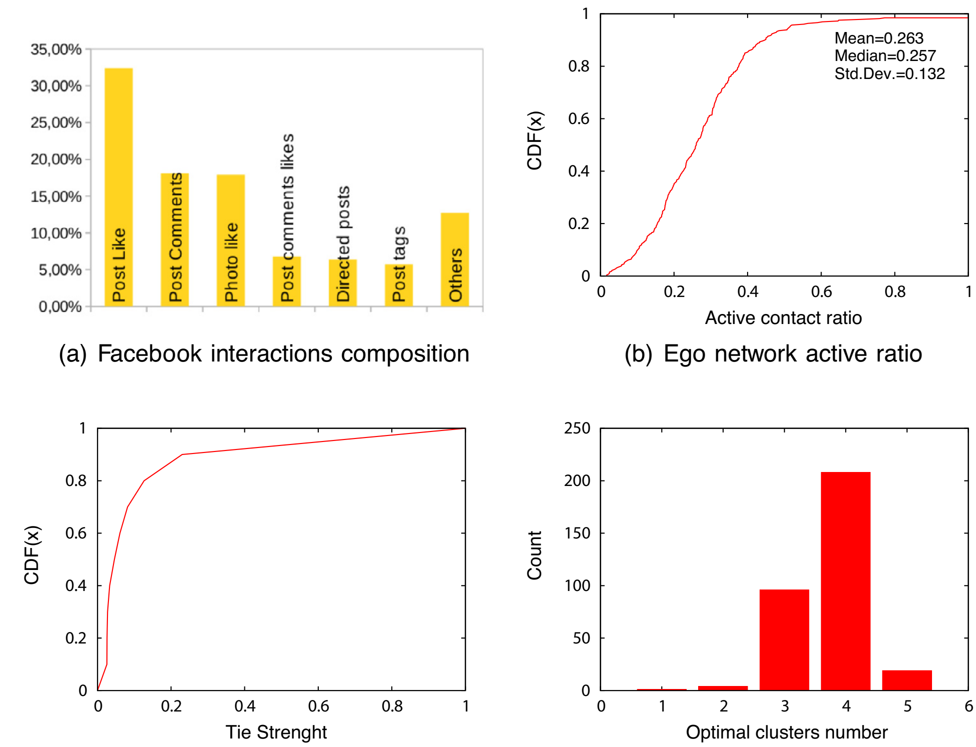
To reflect different levels of importance of the relationships, friendships are associated to a tie strength, a numerical value representing the social distance between the ego and the alter involved

[4] Available at: http://socialcircles.eu/

[5] http://snap.stanford.edu/data/egonets-Facebook.html

Table 1
Descriptive characteristics of the dataset.

Variables	Mean	Median	StdDev	Min	Max
Ego network friends	486.89	394	361.60	23	2965
Ego network friendships	10930.14	4586	22218.14	62	256968
Clustering	0.63	0.63	0.078	0.42	0.88
Communities number	5.81	6	1.62	2	11
Local degree	0.068	0.053	0.053	0.014	0.59
Modularity	0.49	0.51	0.14	0.072	0.766
Ego betweenness centrality	0.724	0.755	0.139	0.07	0.95



(a) Facebook interactions composition (b) Ego network active ratio (c) Tie strength distribution (d) Optimal clusters number

Fig. 1. Analysis of the users interactions and ego network structure.

in a relationship. It has been shown [2,22] how tie strength between egos and their alters is strongly related to their contact frequency, computed as the ratio between the number of direct interactions and the duration of the social relationship. Since both Facebook API are not able to provide the overall duration of a friendship relation and there is a strong correlation between the overall amount of interactions and the contact frequency, we estimate the tie strength as the number of direct interactions occurred from ego to their alter. From our perspective, an active contact is a contact with an associated tie strength greater than 0.

To evaluate the size of active network, we compute for each ego network the ratio between the total number of alters and the number of active alters (*active ratio*), which is a measure of the size of the active network with respect to total network size. The graph in Fig. 1(b) shows that the active ratio of a typical network is around 26%. If we consider the mean size of active network in terms of alters number, we obtain the value of 132, which is comparable to the values discovered by similar studies (i.e., 105 in [22] for OSN, 70.04 in [19] for Twitter and 128.16 in [19] for Facebook).

Finally, we decided to investigate whether females are able to keep more active contacts than males. The analysis of the active network, dividing egos by gender, confirmed this difference: in average women can maintain active connections with 30% of their alters, whereas men only with 23%. In addition, women seem to be more active on OSN than men, with an average activity per alter of 4.8 compared to 3.9 of men.

We try to understand better the tie strength nature by focusing on how the tie strength is distributed among alters: for each ego network, we computed the alters tie strength distribution and built a aggregated CDF shown in Fig. 1(c), where *min-max normalization* has been used for. The elbow in the graph indicates that around 10% of alters can be considered at a high level of intimacy and trust: compared to analysis in [22], which provided a value of 23.53% by considering the recency of contact as tie strength model, we obtain a lower value. We explain this with the higher average size of the ego networks.

We investigate whether the ego networks expose the same concentric structures known as Dunbar circles (as described in Section 2). We perform a mono-dimensional clustering analysis using the K-Means algorithm [26], by exploiting the tie strength value separately for each active network of our sample. To compute the best number of clusters, we adopt the well-known *elbow* method [26] adding a new cluster iteratively until the improvements of the clusters are below the 0.1 threshold. Fig. 1(d) shows the distribution of the optimal number of clusters for each ego network (95% C.I. $\pm$ 0.065). The majority of active networks (63.4%) have an optimal clusters number of 4. As regards the ego networks with 3 clusters, we found that their structure depends on both the lower ratio of user activity and the lower number of social links they expose. In fact, compared to the users with 4

Please cite this article as: A. De Salve et al., The impact of user's availability on on-line ego networks: A Facebook analysis, Computer Communications (2015), http://dx.doi.org/10.1016/j.comcom.2015.09.001

Table 2
Dunbar circles analysis for networks with k-opt = 4.95% confidence intervals are reported in square brackets.

	C_1	C_2	C_3	C_4
Size	4.3 [.50]	17.7 [1.85]	50.8 [5.19]	132.8 [13]
Scal. f.	-	3.64 [.46]	2.35 [.24]	2.69 [.35]
Mean TS	34.09 [2.49]	12.69 [.43]	5.05 [.1]	1.57 [0.017]
Median TS	22	9	4	1
Min TS	23	7	3	1
Max TS	363	149	52	15
Properties of active network				
Network size			416.5	
Outgoing active network perc.			28%	
Ingoing active network perc.			30%	
Total ego interactions			461	
Sex of ego (m/f)			61% / 39%	

Dunbars' circles they have smaller average ego network size (362.5 vs. 416.5), less overall interactions (338 vs. 461), smaller daily session number (4.5 vs. 4.9) and session length (4.84 vs. 5.50 time slots). These facts seem to suggest that ego networks with three clusters are composed by users which don't use Facebook as much as other groups, confirming a claim stated in [2]. Since ego networks with three Dunbar circles do not have a counterpart in real ego networks, we will focus only on ego networks with a number of circles equals to 4. The detailed results about the obtained clusters (or circles) for these ego networks are shown in Table 2. For each circle (C_1, C_2, C_3 and C_4 ordered from the innermost to the outermost), we computed the average size of the Dunbar circle for each ego network (*size*), the scaling factor between circles (*scal.f.*), the minimum/maximum tie strength (*min/max TS*), the mean (*mean TS*) and the median (*median TS*) tie strength of the circles. Finally, we added some properties which characterize globally these networks, such as their average size, their average active network and the average number of total interactions performed by egos.

These networks confirm the Dunbar circles hypothesis: in particular we notice the average scaling factor between the concentric circles sizes is about 3: similar values have been demonstrated to hold in [2] (scaling factor of 3.12) and [19] (scaling factor of 3.45). Compared to [2,19], the Dunbar circles seem to be slightly bigger; however, we highlight that our data comes from complete ego network structure while networks in [2] have been subject to an estimation due to partial topology. We believe that the approximation may underestimate the real circles size. Finally, our results also indicate that the size of the Dunbars' circles is very similar with other results about offline social networks [21].

5. Analysis of the user behaviors

The main stage of our analysis involved the study of users temporal behavior in Facebook. Since there is no direct way to obtain the time spent online by users and their friends, we use the *chat status service* to track the online status of Facebook's users. We sampled all the 337 registered egos and their friends every 8 minutes for 10 consecutive days (from Tuesday 3 June 2014 to Friday 13 June). Using this methodology we were able to access the temporal status of 308 registered users and of their friends (for a total of 95.578 users). For the purpose of clarity, we will refer to registered users to indicate these 308. In order to characterize OSN workloads at the session level, we consider the availability trace of each user to determine the start of a session (when a user switches from *offline* to *online* or *idle*) or the end of a session (when a user switches from *online* or *idle* to *offline*). Utilizing the session information, we first examined the number of concurrent users that accessed the OSN site (see Fig 2(a)). The plot indicates clearly the presence of a cyclic day/night pattern (confirming the results in [18]). Since the majority of the registered users live in Italy or in central Europe, time-zone differences are negligible. The analysis of graph depicts the presence of two peaks: on average, most users seem to be connected after lunch time with a peak around 14:30. The other peak is usually in the evening, around 22:30, probably preceding the sleeping time. It is interesting to notice that the presence of weekend seems to have no influence on users: Friday and Saturday night seem not to expose the above mentioned evening peak, reflecting the fact that many people may go out. It is important to notice that these patterns describe just a global tendency, and cannot be exploited to make any prediction nor assumption of single user behavior.

In order to estimate how often and for how long users connect to OSN, we measure the frequency and duration of sessions for each user. Fig. 2(b) shows how many sessions are done by users (95% C.I. $\pm$ 0.23). We can notice that the majority of users (90%) exposes on average less than 100 daily sessions while the average number of sessions for all users is less than 4 sessions per day.

Fig. 2 (c) shows, for all users, the CDF of the session length (95% C.I. $\pm$ 0.65) and the elapsed time (*inter arrival time*) between two consecutive user's sessions (95% C.I. $\pm$ 4.44). There is a large variation in the OSN usage among users. However, almost half of user sessions are shorter than 20 min (median value of 24 min), and a significant percentage of 34% last less than 10 min. Only a few users sessions (less than 13%) have a long duration, exceeding the 2 h. We can notice that almost 50% of users present an inter-arrival time shorter than 1 hour. These plots confirm therefore the fact that in OSN the typical session has a short duration. Small inter-arrival times correspond to users who constantly use the OSN service, while large inter-arrival times correspond to users who connect occasionally to the OSN. It is important to notice that the size of the active network of each user is slightly correlated to the time spent online by the ego, such as the average session length (0.36) and the number of daily sessions (0.20): intuitively we would expect that ego with more active contacts are likely to spend more time on the OSN.

While researchers have seen that strong and weak ties are characterized by different levels of *homophily*, i.e. the tendency of individuals with similar interests to join with each other, it has not been understood to what extent ties in circles show the existence of *temporal homophily*, the tendency of similar individuals to participate in similar uptime patterns. Moreover, the actual impact of correlated availabilities on Dunbar circles remains unexplored. In order to bridge this gap, we evaluate whether the online patterns of users are correlated with those of their alters in each Dunbar circles. We consider separately the alters in each circle and compute the availability correlation between egos and their alters using the similarity between their availability patterns. As done in [5], we evaluate this correlation using the *cosine similarity* metrics [27] which is frequently adopted when trying to determine similarity between binary data (such as documents or in our case availability patterns). The availability of each user is represented by an availability vector of fixed size. For each time slot of the monitoring period (eight-minute time slot for 10 consecutive days) the corresponding entry contains 1 if the user was online at that time and 0 otherwise. Formally, let A and B the availability vector of two users, the cosine similarity is computed as shown in Eq. (1):

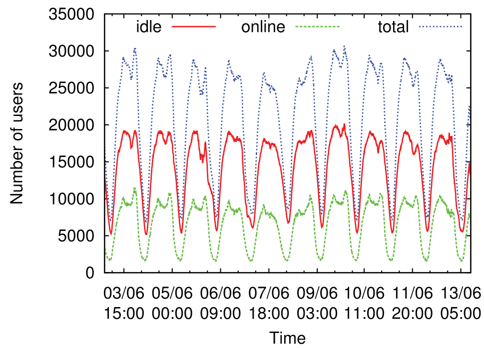
$$CosineSimilarity(A, B) = \frac{A \cdot B}{||A|| \cdot ||B||} \quad (1)$$

The resulting similarity ranges from 1 meaning perfect correlation, to 0, usually indicating no correlation between the ego and the alter.

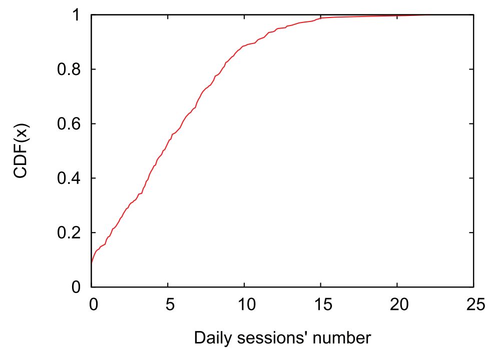
We investigate to what extent the availability patterns of friends who appear to be online in the same time slot are similar, for different time windows of the day. We divide the day into 6 time windows of four hours each (from 0 to 24), and then compute the cosine similarity of the availability patterns among friends who appear to be online at the same time slot of the considered time window. The Fig. 2(d)

Please cite this article as: A. De Salve et al., The impact of user's availability on on-line ego networks: A Facebook analysis, Computer Communications (2015), http://dx.doi.org/10.1016/j.comcom.2015.09.001

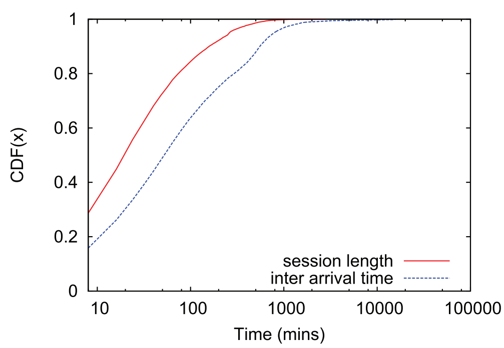
ARTICLE IN PRESS
JID: COMCOM [m5G;October 14, 2015;12:1]



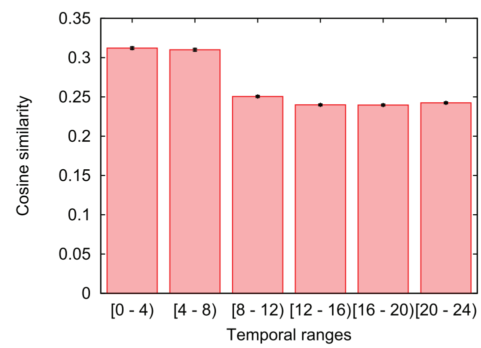
(a) Availability of all users



(b) Number of sessions for all users



(c) Sessions length and inter arrival time for all users



(d) Similarity of the online-offline patterns on temporal ranges

Fig. 2. Analysis of the temporal properties.

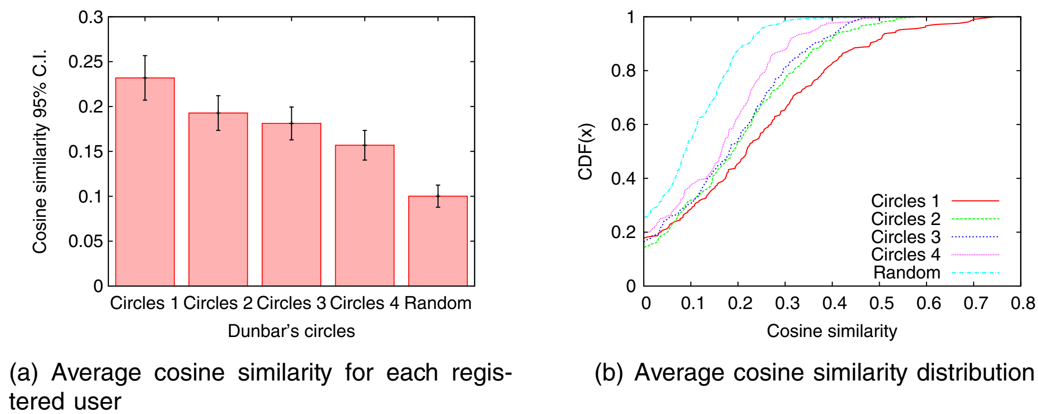
shows the average similarity for each time window. As we expected, users who happened to be both online during abnormal temporal window (from 0am to 8am) have greater similarity than users connected during the classical time windows (from 8am to 0am, i.e. preceding the sleeping time).

Since a correlation implicitly exists at the dataset level as users of the same country tend to connect during specific times of the day (see Fig. 2(a)), we compare availability correlation on each circle with those obtained by considering external users. For these purposes, we compute also the average correlation between the ego and the set of their friends outside the Dunbar's circles (referred as Random). We have computed, for each registered user, the average cosine similarity between the egos and their alters in each of the Dunbar's circles. Fig. 3(b) shows the CDF of correlation values, while Fig. 3(a) shows the average correlation values for each circle. The similarity values are rather low for the Dunbar's circles as well as for random friends. However, the graph clearly indicates that alters in innermost circles (such as *Circles 1* and *Circles 2*) have a higher average similarity with the availability pattern of the ego than alters in outermost circles (such as *Circles 3* and *Circles 4*). The similarities with alters in Dunbar's circles are higher than with random ones, thus highlighting the impact of Dunbar's circles on availability. The average similarity of each circle is equal respectively to: 0.23 for Circle 1 (95% C.I. ± 0.025) , 0.19 for Circle 2 (95% C.I. ± 0.019), 0.18 for Circle 3 (95% C.I. ± 0.018), 0.15 for Circle 4 (95% C.I. ± 0.017) and 0.10 for random friends (95% C.I. ± 0.012).

We have also measured the average number of times ego and their alters are both online (1,1) or offline (0,0) in the same time slot and the number of times that either ego (1,0) or alter (0,1) are online, separately for each circle. The Fig. 4(a), (b), (c) and (d) show the CDF of the (0,0), (0,1), (1,0) and (1,1) matching, respectively. As we expected, the number of (1,1) matching between availability vectors increases as we consider alters of the innermost circles. In contrast, the number of (0,0) matching indicates an opposite trend since they decrease as much as we consider alters the inner circle. This highlights the key role that positive matches of the form (1,1) have on the availability pattern of close alters compared to the matches of the form (0,0). As regards the (0,1) matching results show that, when the ego is offline, alters in the inner circles are much more online than alters belonging to the outer circles. An opposite trend occurs when we consider the (1,0) matching since the number of offline alters when ego is online appear to increase as we consider outer circles. In order to characterize the temporal structure of the active network, we compute the average percentage of matching found in each circle (see Fig. 4(e)). In order to test if the results are independently of the way we define the active network, we have computed the tie strength of each user by considering: *(i)* the interactions over the whole crawling period (6 Months), *(ii)* the interactions occurred in the last three months of the crawling period (3 Months) and *(iii)* the interactions occurred in the last month (1 Months). The average percentages of matching resulting after this filtering remain quite similar and do not depends very much on the time window we use to define an active contact.

The average percentage of (0,0) matching exceeds other cases in each circles and it is equal to: 61.4% for Circle 1 (95% C.I. ± 8.4), 62.4% for Circle 2 (95% C.I. ± 7.3), 63.5% for Circle 3 (95% C.I. ± 7.2), 65.4% for Circle 4 (95% C.I. ± 7.3) and 67.6% (95% C.I. ± 7.4) for alters who are not members of the Dunbar's circles. Instead, matching of the form (1,1) has the lowest percentage value for all circles, namely: 7.4% for Circle 1 (95% C.I. ± 3.9), 5.9% for Circle 2 (95% C.I. ± 2.6), 5.3% for Circle 3 (95% C.I. ± 2.2), 4.6% for Circle 4 (95% C.I. ± 2.0) and 3.6% (95% C.I. ± 1.6) for random friends. The Circle 1 has approximately the same percentage of (0,1) and (1,0) matching, i.e. 15.9% (95% C.I. ± 4.5) and 15.3% (95% C.I. ± 5.6), respectively. As alters are located in the outer circles the percentage of the (1,0)/(0,1) matching increases/decreases

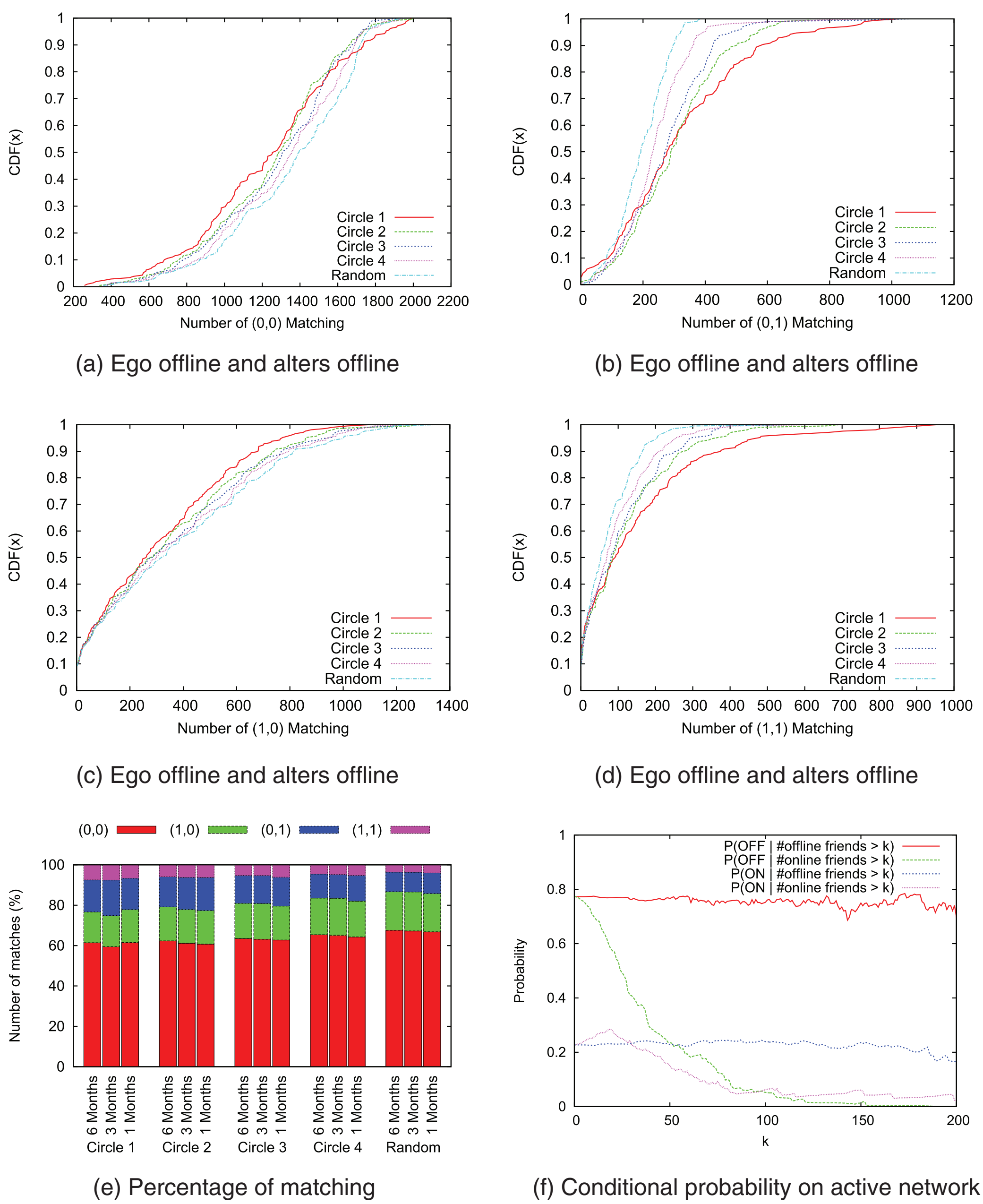
Please cite this article as: A. De Salve et al., The impact of user's availability on on-line ego networks: A Facebook analysis, Computer Communications (2015), http://dx.doi.org/10.1016/j.comcom.2015.09.001



(a) Average cosine similarity for each registered user

(b) Average cosine similarity distribution

Fig. 3. Analysis of the Dunbar circles temporal features.



(a) Ego offline and alters offline

(b) Ego offline and alters offline

(c) Ego offline and alters offline

(d) Ego offline and alters offline

(e) Percentage of matching

(f) Conditional probability on active network

Fig. 4. Analysis of the temporal matching on Dunbar circles.

of about 1%. As a further step, we characterized the impact of this similarity between users and their friends on the probability of the ego to be online/offline by taking into account the aggregated behavior of their alters in the active network. In order to estimate this impact, we computed for each ego the probability to be online/offline depending on the available alters in each time slot. As done in other studies [5,15], we model this dependence using conditional probabilities. More formally, let $e = \{ON, OFF\}$ the events "ego is online/offline" and $a_k = \{\#\ online/offline\ friends > k\}$ the set of event "At least k alters of the active network are online/offline", we calculated the conditional probability $P(e|a_k) = P(e \cap a_k)/P(a_k)$ for $k = 0, \ldots, 200$. The value $P(e \cap a_k)$ is the number of time slots when the user is online/offline and at least k of her active contacts are online/offline, normalized by the total number of time slots. $P(a_k)$ is the number of time slots when at least k of her active contacts are online/offline, normalized by the total number of time slots. Fig. 4(f) shows the CDF of the conditional probabilities for different combinations of events. The results clearly show that an ego is more likely to be online when at least 10 of their alters in Dunbar's circles are connected. After that, the conditional probability decreases as the number of online friends increases. The probability that ego is offline decreases very quickly when the number of the online neighbours increase. Instead, the conditional probability that the ego is offline/online remains roughly the same, for any number of offline neighbours.

6. Conclusion

Our study uncovered a number of interesting findings related to the specific nature of online social networking environments. By using temporal information of real OSNs we have found that availability patterns of single individuals has a non-trivial relationships with those of their close friends. We found that the size of the active network of each user is slightly correlated to the time spent online by the ego, such as the average session length (0.36) and the number of daily sessions (0.20). We have shown the extent to which availability patterns of each Dunbar's circle affect the availability pattern of the user. Namely, social ties on innermost circles not only are stronger in terms of volume of communications, but also show higher similarity of the users availability pattern. We have also measured the average number of times ego and their alters are both online (1,1) or offline (0,0) in the same time slot and the number of times that either ego (1,0) or alter (0,1) are online, separately for each circle. The average percentage of (0,0) matching increases as we consider alters located in the outermost circles. The matching of the form (1,1) has the lowest percentage value for all circles and they increases as we consider users of the innermost circles. Finally, the innermost circle has approximately the same percentage of (0,1) and (1,0) while for the outer circles, the percentage of the (1,0)/(0,1) matching increases/decreases of about 1%. By characterizing the impact of this similarity between the availability pattern of the users and their friends in term of conditional probability, we have shown that users have more probability to be online when at least 10 of their Dunbar's friends are online . Finally, we have shown that users who happened to be both online during abnormal temporal window (from 0am to 8am) have greater similarity than users connected during the classical time windows (from 8am to 0am, i.e. preceding the sleeping time). These understanding have a significant impact on the success/failure of the OSN and they could lead to many benefits if properly considered and addressed during the design of the designing of the next-generation OSN infrastructures. As future work, we would like to investigate the impact of our finding on content distribution patterns. Answering these questions will let us explore opportunities for efficient content distribution and data replication, as well as advertisement and recommendation strategies. Lastly, based on our results, we plan to build a user churn model able to shape availability pattern of the user behavior by incorporating the most part of our findings, including sessions distribution, tie strength and temporal features.

References

[1] V. Arnaboldi, M. Conti, A. Passarella, F. Pezzoni, Ego networks in twitter: an experimental analysis, in: INFOCOM, 2013 Proceedings IEEE, IEEE, 2013, pp. 3459–3464.
[2] V. Arnaboldi, M. Conti, A. Passarella, F. Pezzoni, Analysis of ego network structure in online social networks, in: Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom), IEEE, 2012, pp. 31–40.
[3] E. Gilbert, K. Karahalios, Predicting tie strength with social media, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2009, pp. 211–220.
[4] L. Jin, Y. Chen, T. Wang, P. Hui, A.V. Vasilakos, Understanding user behavior in online social networks: A survey, IEEE Commun. Mag. 51 (9) (2013) 144–150.
[5] A. Boutet, A.M. Kermarrec, E. Le Merrer, A. Van Kempen, On the impact of users availability in OSNs, in: Proceedings of the Fifth Workshop on Social Network Systems, ACM, 2012, p. 4.
[6] D. Schiöberg, F. Schneider, G. Trédan, S. Uhlig, A. Feldmann, Revisiting content availability in distributed online social networks, CoRR (2012) arXiv:1210.1394.
[7] M. Conti, A.D. Salve, B. Guidi, F. Pitto, L. Ricci, Trusted dynamic storage for dunbar-based P2P online social networks, in: Proceedings On the Move to Meaningful Internet Systems: OTM 2014 Conferences - Confederated International Conferences: CoopIS and ODBASE 2014, 2014a, pp. 400–417.
[8] M. Conti, A.D. Salve, B. Guidi, L. Ricci, Epidemic diffusion of social updates in dunbar based DOSN, in: Second Workshop on Large Scale Distributed Virtual Environments on Clouds and P2P, LSDVE 2014, Euro-Par 2014: Parallel Processing - 20th International Conference - Lecture Notes in Computer Science, 8632, 2014b.
[9] R.E. Wilson, S.D. Gosling, L.T. Graham, A review of facebook research in the social sciences, Perspect. Psychol. Sci. 7 (3) (2012) 203–220.
[10] G. Greenwald, E. MacAskill, Nsa prism program taps in to user data of apple, google and others, Guardian 7 (6) (2013) 1–43.
[11] T. Paul, A. Famulari, T. Strufe, A survey on decentralized online social networks, Comput. Netw. 75 (2014) 437–452.
[12] M.S. Granovetter, The strength of weak ties, Am. J. Soc. (1973) 1360–1380.
[13] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. 2008 (10) (2008) P10008.
[14] L. Gyarmati, T.A. Trinh, Measuring user behavior in online social networks, Netw. IEEE 24 (5) (2010) 26–31.
[15] R. Bhagwan, S. Savage, G.M. Voelker, Understanding availability, in: Peer-to-Peer Systems II, Springer, 2003, pp. 256–267.
[16] J.J. Jones, J.E. Settle, R.M. Bond, C.J. Fariss, C. Marlow, J.H. Fowler, Inferring tie strength from online directed behavior, PloS one 8 (1) (2013) e52168.
[17] F. Benevenuto, T. Rodrigues, M. Cha, V. Almeida, Characterizing user behavior in online social networks, in: Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference, ACM, 2009, pp. 49–62.
[18] S.A. Golder, D.M. Wilkinson, B.A. Huberman, Rhythms of social interaction: Messaging within a massive online network, in: Communities and Technologies 2007, Springer, 2007, pp. 41–66.
[19] B. Gonçalves, N. Perra, A. Vespignani, Modeling users' activity on twitter networks: Validation of dunbar's number, PloS one 6 (8) (2011) e22656.
[20] R.I. Dunbar, The social brain hypothesis, Brain 9 (10) (1998) 178–190.
[21] A. Sutcliffe, R. Dunbar, J. Binder, H. Arrow, Relationships and the social brain: Integrating psychological and evolutionary perspectives, Br. J. Psychol. 103 (2) (2012) 149–168.
[22] V. Arnaboldi, A. Guazzini, A. Passarella, Egocentric online social networks: Analysis of key features and prediction of tie strength in facebook, Comput. Commun. 36 (10) (2013) 1130–1144.
[23] C. Wilson, A. Sala, K.P. Puttaswamy, B.Y. Zhao, Beyond social graphs: User interactions in online social networks and their implications, ACM Trans. Web (TWEB) 6 (4) (2012) 17.
[24] R. Gracia-Tinedo, M. Sanchez Artigas, P. Garda Lopez, Analysis of data availability in f2f storage systems: When correlations matter, in: Peer-to-Peer Computing (P2P), 2012 IEEE 12th International Conference on, IEEE, 2012, pp. 225–236.
[25] M. Everett, S.P. Borgatti, Ego network betweenness, Soc. Netw. 27 (1) (2005) 31–38.
[26] D.J. Ketchen, C.L. Shook, The application of cluster analysis in strategic management research: an analysis and critique, Strat. Manag. J. 17 (6) (1996) 441–458.
[27] S. seok Choi, S. hyuk Cha, A survey of binary similarity and distance measures, J. Syst. Cybern. Inf. (2010) 43–48.

Please cite this article as: A. De Salve et al., The impact of user's availability on on-line ego networks: A Facebook analysis, Computer Communications (2015), http://dx.doi.org/10.1016/j.comcom.2015.09.001