

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

^aUniversity of Pisa, Department of Computer Science, Largo Bruno Pontecorvo, Pisa, Italy

^bInstitute for Informatics and Telematics, IIT, CNR, Via Moruzzi, Pisa

Abstract

Online Social Networks (OSNs) are the most popular applications in today's 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 online social networks 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 organised into a set of circles (named *Dunbar's circles*) according to the *ego network model*. In this model, circles are arranged into a hierarchical inclusive sequence based on an increasing level of intimacy. 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 temporal properties that characterize an ego network.

Keywords: Online Social Networks, temporal homophily, Dunbar's circles

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 behaviours [1, 2, 3], computer scientists and engineers are focusing on the technical side by studying their properties [4, 5], improving these architectures [6, 7, 8, 9] and optimizing their usage [10].

Availability of OSNs datasets has led to better understanding of OSNs at global structural level [11, 12, 13], as well as at the level of local structural properties [2, 1, 14, 13]. Since few years, increased awareness about the importance of temporal information has led scientists to investigate the temporal behaviour of users in OSNs: the study of online sessions will help in characterizing the typical OSNs' usage and understanding how users interact with these platforms. In the last few years, the study of ego networks [15] in virtual environments

has received more and more attention. Ego networks are social networks made up of an individual (*ego*) and all the social relations it has with other people (*alters*). Despite numerous studies, some characteristics of ego networks still remain unexplored, particularly their temporal characteristics..

In general, temporal properties of OSNs' users are important for a variety of purposes. First, studies of user behaviours allow the performance of existing systems to be evaluated and lead to better site design. Second, understanding ego network temporal properties in OSNs is crucial for social studies as well as for the study of the dynamic processes that occur on them (such as 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 [7] and Content Distribution Systems [8, 9, 6].

While a few recent studies examined the availability patterns of OSNs' users in terms of session length or interactions frequency, they do not provide a global picture of how temporal properties of users affect their ego networks. The presence of a huge amount of records containing users' information, provided by OSNs, represents new opportunity to analyse and better understand the ego network temporal properties. The definition of tools and analysis exploiting this data is an urgent need since public available data regarding both social relationships and behaviour of users is getting more and more difficult to be obtained from OSNs. Nowadays, many popular OSNs have started to restrict their privacy policies to limit the amount of user's data accessed by third-party applications.

The aim of this paper is to investigate whether it is possible to find the presence of temporal dependence in 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 ¹. Using this sample of Facebook users, we find that ego network structure have a significant impact on temporal availability patterns of the ego. We identify an increasing level of similarity (or *temporal homophily*) between availability patterns of the ego and alters in their ego network.

The remainder of this paper is organised as follows. Section 2 describes the basic concepts used and the related work in the contexts of: user behaviours and structural properties in OSNs. 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 Section 4.3), the interactions of the users (Section 4.2) and the their temporal features (Section 4.4). In this section we also discuss our results and compare them with other past studies. Finally, Section 5 draws the main conclusions of these analyses.

¹Available at: <http://socialcircles.eu/>

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 OSNs and (ii) existing studies reporting the analyses of the user behaviours in OSNs.

Analyses of OSN structural properties. 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 [10]. 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 [2, 1, 16], and they found that structures similar to those of off-line social networks can also be identified in OSNs. As shown by Dunbar [17], there are constraints limiting the number of active relationships ego 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 [15]. As discovered by Dunbar [17], ego networks are organized into a hierarchical structure formed of 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 [18, 2, 16, 1]. The design of a measures of tie strength is still an open issue and some recent work try to build a model for tie strength computation using data coming from OSNs [3, 14, 19]. 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 [14, 2, 1] use frequency of the online interactions to estimate tie strength, since the two concepts appear to be tightly correlated. Finally, authors in [19] showed that, while frequency of interaction is a necessary component for predicting the strength of a tie, additional information related to social interactions is helpful to achieve more accurate predictions.

Analyses of OSN user behaviours. The work in [20, 5, 21, 12, 4] studied the user behaviours by focusing on temporal properties. Golder et al. [20] showed 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] used traces relative to MySpace OSN to study the correlation between each pair of users availability traces. They observed that users are more present when

their top friends are online rather than when their random friends are online. Authors in [21] analysed the workloads of the users who accessed four popular social networks. They studied 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 et al. [12] analysed user behaviour in OSNs (namely Bebo, MySpace, Netlog and Tagget). They observed 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] conducted a comprehensive review about the user behaviour in OSNs from several perspectives, including connectivity and interaction among users. Finally, several recent works [6, 7, 22] analyse the availability of the user to provide efficient storage strategies and data persistence in DOSN. Authors in [22] evaluated several storage and replication strategies in a Friend-to-Friend (F2F) storage system and they found that availability correlations offers a good trade-off between data availability and data redundancy. Authors in [6] exploited 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, for the best of our knowledge, none of the studies above provides a detailed analysis of the temporal properties 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 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 common to work with small but representative sample of an OSN. As far as concerns the study of the users' behaviour, most studies utilize synthetic users' availability traces, and in the context of distributed systems, some studies used synthetic availability traces provided by P2P well-known churn models. However, the scientists realized that these traces were not useful since nature of the P2P systems compared to OSN is very different [13]. At the best of our knowledge, no existing updated dataset is able to provide complete information, such as information regarding the social graph, interactions among users and temporal information (online sessions) for a real OSN.

Furthermore, public available data regarding both social relationships and behaviour of users is getting more and more difficult to be obtained from OSNs. Nowadays, many popular OSNs have started to restrict their privacy policies to limit the amount of user's data accessed by third-party applications. For instance, since 1st May 2014 Facebook introduced the new API which removed

the permission granted to applications to access some users data, such as the only accessible temporal information: the online chat status.

For these reasons we have implemented a Facebook application, called *SocialCircles!*², 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 interaction 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 *offline*, 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 minutes).

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). Since few users de-authorized our application, we were able to retrieve complete interactions information of 328 users. The resulting Facebook population have 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 thus expect less biased results and more variety in the networks structure, which should better reflect real OSN utilization. Using *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 temporal properties which characterize the Dunbar ego networks in Facebook and *iv)* analyse the extent to which the availability of and ego depends on the presence of the alters in each Dunbar's circles.

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

²Available at: <http://socialcircles.eu/>

found in past research. We also investigate the relationship existing between ego network structures and the availability of a user in the system.

4.1. Social Graph analysis

Initially, we analysed the collected data from a pure topological perspective. For each ego network, we have collected some measures such as the number of nodes, number of edges and the average clustering coefficient. This analysis has been performed on the whole set of 337 ego networks. Fig. 1(a) shows the *Cumulative Distribution Function (CDF)* of friends for all the analysed ego networks (95% C.I. \pm 38.75). We notice that the majority (80%) of ego networks have less than 600 friends, whereas only 7% of nodes exceed 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 Fig. 1(e) shows the CDF of friendships between ego and alters and between alters, i.e the total number of edges in the ego network (95% C.I. \pm 2380.72). To better show the results, a logarithmic scale has been used. 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 network 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 analysed sample. indicates therefore the presence of a tightly connected graph structure. Fig. 1(b) shows the CDF of the clustering coefficient for our sample (95% C.I. \pm 0.008). A high average clustering coefficient equal to 0.636 indicates the presence of a tightly connected graph structure. We notice that the sample analysed exposes a comparable but slightly higher average clustering value with respect to past analyses (such as 0.6055³). We evaluated the local degrees distribution of nodes: a centrality index that measures the number of mutual friends that an alter shares with each ego. Fig. 1(d) shows the CDF of the local degree between the ego and his alter. Specifically, ego network and alter shares on average 4% of friends with the ego.

To achieve a deeper understanding of crawled ego networks we performed a clustering analysis to discover topology-based communities. We applied the Louvain community detection method [11] that aims to maximizing the modularity value assigned to each node: a measure of how much a network is composed by several communities of nodes. Fig. 1(f) shows how the modularity value is distributed among all ego networks (95% C.I. \pm 0.013), whereas Fig. 1(c) shows the result of the partitioning, displaying the number of discovered communities in each ego network (95% C.I. \pm 0.17). This analysis confirms that the ego networks are composed by different communities.

³<http://snap.stanford.edu/data/egonets-Facebook.html>

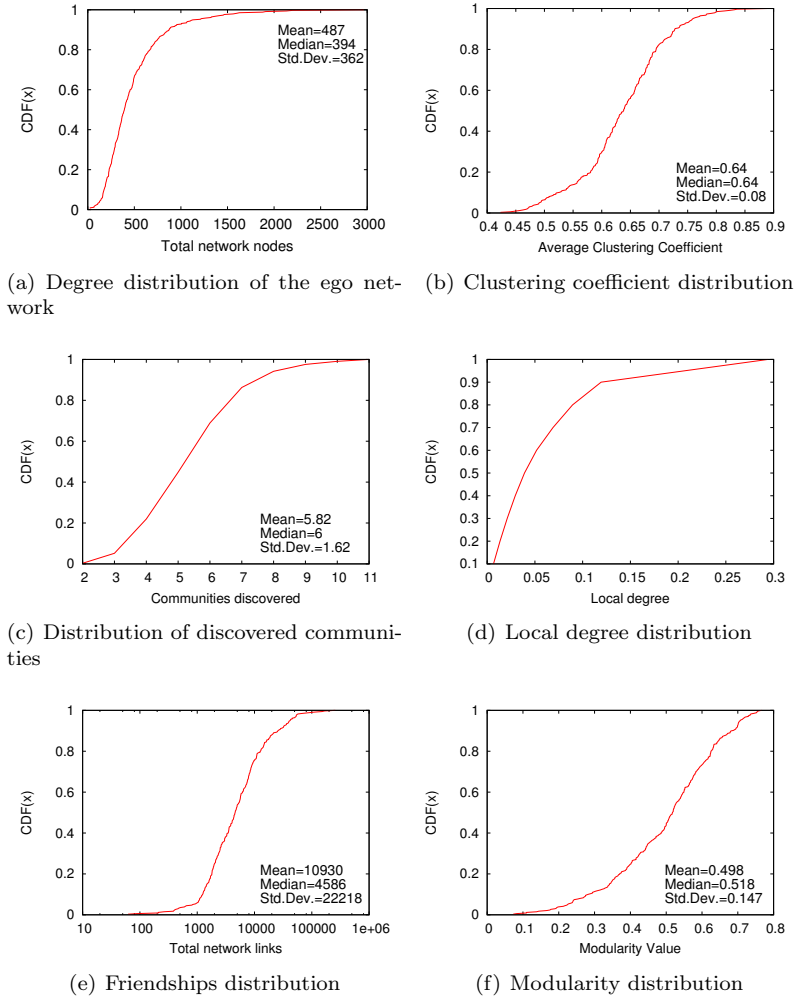


Figure 1: Analysis of the OSN properties

4.2. Analysis of the user interaction

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 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. 2(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 [19]. 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.

By considering both direction and amount of visible interactions we are able to estimate the importance of the tie between ego and alters (tie strength). As already discussed in different studies [19, 2], tie strength between ego and their alter 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 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 within 6 months before the end time of the crawler. In 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 Figure 2(b) shows that the active ratio central tendency 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 117.8, which is comparable to the values discovered by similar studies (e.g., 105 in [19] for OSN, 70.04 in [16] for Twitter and 128.16 in [16] 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 seems to be overall 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 (normalized) alters tie strength distribution and build an aggregated CDF shown in Fig. 2(c). 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 [19], 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.

4.3. Analysis of the ego network properties

In this section we investigate whether the ego networks expose the same concentric structures known as Dunbar circles (as described in Section 2). We

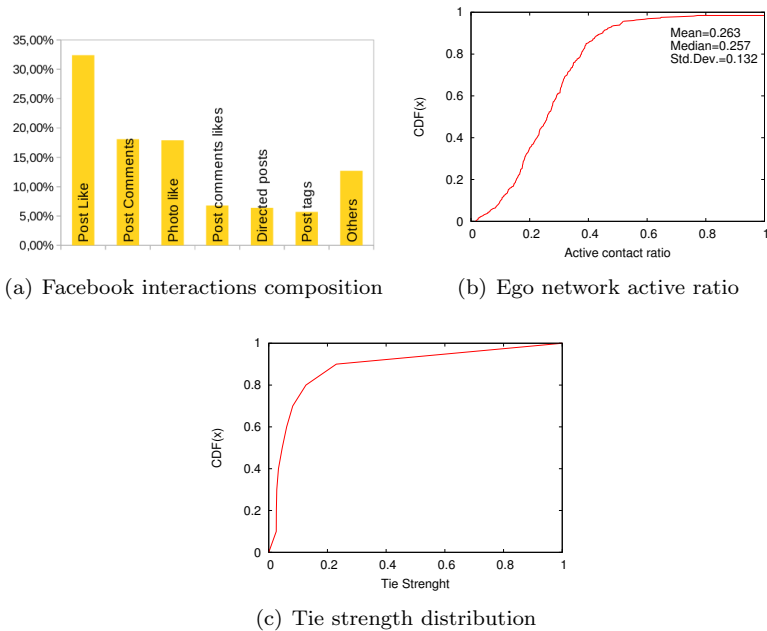


Figure 2: Analysis of the users interactions

perform a mono-dimensional clustering analysis using K-Means [23], 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 [23] adding a new cluster iteratively until the improvements of the clusters is below the 0.1 threshold. Figure 3 shows the distribution of the optimal number of clusters for each ego network (95% C.I. ± 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 suppose that their structure depends on both the lower ratio of user activity and the lower number of social links they expose. These facts seem to suggest that ego networks with three clusters are composed by users which probably don't use Facebook as much as other groups, exposing an overall minor usage (confirming a claim stated in [2]) and perhaps still giving more importance to other communication channels. 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 equal to 4. The detailed results about the obtained clusters (or circles) for these ego networks are shown in Table 1. For each circle (C_1 , C_2 , C_3 and C_4 ordered from the innermost to the outermost) this table indicates the average size of the Dunbar circle for each ego network (*size*), the scaling factor between circles (*scal.f.*) and the minimum tie strength of the circle (*min TS*). 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 which is equal to 3.2: similar values have been demonstrated to hold for OSN [2]. Compared to [2], our discovered Dunbar circles seem to be slightly bigger: we highlight however that our data comes from complete ego networks 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.

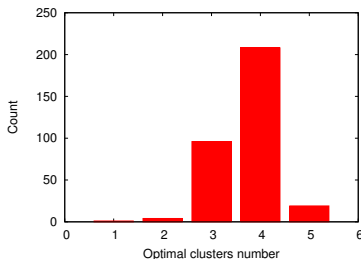


Figure 3: Optimal clusters number

	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.	-	4.1	2.9	2.6
min TS	23	7	3	1
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%			

Table 1: Dunbar circles analysis for networks with $k\text{-opt} = 4$. 95% confidence intervals are reported in square brackets.

4.4. Analysis of the user behaviours

The last stage of our analysis involved the study of users temporal behaviour 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 308 registered users and the set of their friends (for a total of 95.578 users). For the purpose of clarity, we will refer in this section 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 4(a)). The plot indicates clearly the presence of a cyclic day/night pattern (confirming the results in [20]). 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 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 behaviour.

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. 4(b) shows how many sessions are done by users (95% C.I. ± 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. 4(c) shows for all users the CDF of the sessions length (95% C.I. ± 0.65) and the elapsed time (*inter arrival time*) between two consecutive user's sessions (95% C.I. ± 4.44). There is a large variation in the OSN usage among users. However, almost half of users sessions are shorter than 20 minutes (median value of 24 mins), and a significant percentage of 34% last less than 10 minutes. Only few users sessions (less than 13%) have a long duration, exceeding the 2 hours. 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-arrivals times correspond to users who constantly use the OSN service while large inter-arrival times correspond to users who connect occasionally to the OSN.

Another very important aspect for OSNs is the role of *weak* and *strong ties*. Weak ties are characterized by lower tie strength and occasional interactions while strong ties correspond to relationships with trusted friends or family. The heterogeneity of social ties is a key ingredient of OSNs and plays a crucial role on information diffusion processes, social influence, information communication, storage and data availability. 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 [24] which is frequently adopted when trying to determine similarity

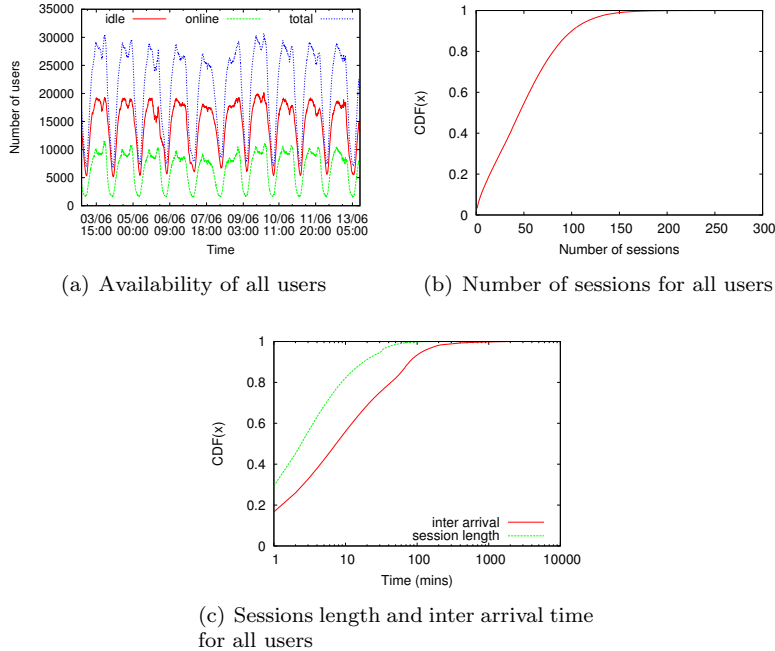


Figure 4: Analysis of the temporal properties

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 timeslot of the monitoring period (eight-minute timeslot 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):

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

The resulting similarity ranges from 1 meaning perfect correlation, to 0, usually indicating no correlation between the ego and alter. 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. 4(a)), we compare availability correlation on each circles 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 users, the average cosine similarity between ego and their alters in each Dunbar’s circles. Fig. 5(b) shows the CDF of correlation values while Fig. 5(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* *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: 68.4 for Circle 1 (95% C.I. \pm 0.025) , 0.19 for Circle 2 (95% C.I. \pm 0.019), 0.18 for Circle 3 (95% C.I. \pm 0.018), 0.15 for Circle 4 (95% C.I. \pm 0.017) and 0.10 for random friends (95% C.I. \pm 0.012). We have also measured the average number of times ego and their alters are both present (1,1) or absent (0,0) in the same timeslots (*positive matches*), separately for each circle (also known as *hamming distance* [24]). We would expect that the number of positive matches between availability vectors increases as we consider alters of the innermost circles. However, results indicate an opposite trend since the average percentage of positive matches decreases for alters in innermost circles: 69.5% for Circle 1 (95% C.I. \pm 18.6) , 70% for Circle 2 (95% C.I. \pm 18), 70.9% for Circle 3 (95% C.I. \pm 19.2), 72.7 for Circle 4 (95% C.I. \pm 20.5) and 75% for random friends (95% C.I. \pm 23.3). This highlights the key role that positive matches of the form (1,1) have on availability pattern of close alter compared to positive matches of the form (0,0).

As a further step, we characterized the impact of this similarity between users and their friends on the probability for users to be connected by taking into account the aggregated behaviour of their alters in each Dunbar’s circle. In order to estimate this impact, we computed for each ego the probability to be connected depending on the available alters, separately for each Dunbar’s circle. As done in other studies [5, 13], we model this dependence using conditional probabilities. More formally, let *ON* the event "ego is online", and *AtLeast50%_k* the set of event "At least 50% of alters in circle *k* are online". We calculated the conditional probability $P(ON|AtLeast50\%_k)$ for each Dunbar’s circle of our sample. Fig. 5(d) shows the CDF of all conditional probabilities while Fig. 5(c) shows the average conditional probability separately for each Dunbar’s circle. As in previous case, probabilities are rather low. However, results clearly show that an ego is more likely to be online when at least half of their alters in Dunbar’s circles are. The average probability is equal respectively to: 0.15 for Circle 1 (95% C.I. \pm 0.028) , 0.025 for Circle 2 (95% C.I. \pm 0.008), 0.005 for Circle 3 (95% C.I. \pm 0.003), 0.001 for Circle 4 (95% C.I. \pm 0.002). We have also computed the average conditional probabilities for the set of friends outside the Dunbar’s circles of the user, resulting in a very low value (0.00002 with 95% C.I. \pm 0.0002). Finally, results show how a clear overlap between the third circle (*affinity group*) and other circles adjacent to it (*sympathy group* and *active network*) did not permit us to characterize properties of the third circle. Each circle of the ego network has distinct characteristics but there is no accurate information currently available in literature about properties of the affinity group. [19, 2].

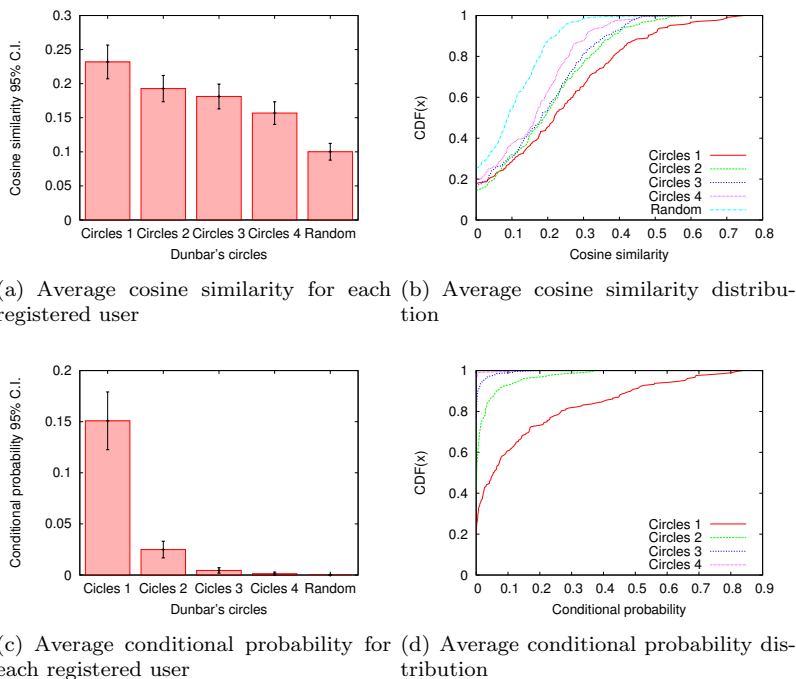


Figure 5: Analysis of the Dunbar circles temporal features

5. 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. Furthermore, we showed the extent to which availability patterns of each Dunbar's circle affect user behaviour. Namely, social ties on innermost circles not only are stronger in terms of volume of communications, but they are also stronger in time. The properties of social relationships between users fundamentally impact the behaviour of the users and can be considered when designing future social networking solutions (such as DOSN). In particular, the study that we conducted adds great value to the several social networking solutions that need to deal with both trust of social ties (such as tie strength) and temporal properties in OSNs, by allowing a better understand and a more solid design and test of future solutions and algorithms based on users' behaviour in OSNs.

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 replicatin, as well as advertisement and recommendation strategies. Lastly, based on our results, we

plan to build a user churn generator able to model user behavior by incorporating the most part of our findings, including sessions distribution, tie strength and temporal features in order to .

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 Conference 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 Communications Magazine* 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. Tredan, S. Uhlig, A. Feldmann, Revisiting content availability in distributed online social networks, arXiv preprint arXiv:1210.1394.
- [7] S. Fu, L. He, X. Liao, K. Li, C. Huang, Analyzing the impact of storage shortage on data availability in decentralized online social networks, *The Scientific World Journal* 2014 (2014) 14.
- [8] 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, Amantea, Italy, October 27-31, 2014, 2014, pp. 400–417.
- [9] Epidemic diffusion of social updates in dunbar based dosn, in: On Second Workshop on Large Scale Distributed Virtual Environments on Clouds and P2P, LSDVE 2014, Euro-Par 2014: Parallel Processing - 20th International Conference, Porto, Portugal, August 25-29 - Lecture Notes in Computer Science, Vol. 8632, 2014, p. 828.
- [10] R. E. Wilson, S. D. Gosling, L. T. Graham, A review of facebook research in the social sciences, *Perspectives on Psychological Science* 7 (3) (2012) 203–220.

- [11] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment* 2008 (10) (2008) P10008.
- [12] L. Gyarmati, T. A. Trinh, Measuring user behavior in online social networks, *Network, IEEE* 24 (5) (2010) 26–31.
- [13] R. Bhagwan, S. Savage, G. M. Voelker, Understanding availability, in: *Peer-to-Peer Systems II*, Springer, 2003, pp. 256–267.
- [14] 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.
- [15] M. S. Granovetter, The strength of weak ties, *American journal of sociology* (1973) 1360–1380.
- [16] B. Gonçalves, N. Perra, A. Vespignani, Modeling users’ activity on twitter networks: Validation of dunbar’s number, *PloS one* 6 (8) (2011) e22656.
- [17] R. I. Dunbar, The social brain hypothesis, *brain* 9 (10) (1998) 178–190.
- [18] A. Sutcliffe, R. Dunbar, J. Binder, H. Arrow, Relationships and the social brain: Integrating psychological and evolutionary perspectives, *British journal of psychology* 103 (2) (2012) 149–168.
- [19] V. Arnaboldi, A. Guazzini, A. Passarella, Egocentric online social networks: Analysis of key features and prediction of tie strength in facebook, *Computer Communications* 36 (10) (2013) 1130–1144.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] D. J. Ketchen, C. L. Shook, The application of cluster analysis in strategic management research: an analysis and critique, *Strategic management journal* 17 (6) (1996) 441–458.
- [24] S. seok Choi, S. hyuk Cha, A survey of binary similarity and distance measures, *Journal of Systemics, Cybernetics and Informatics* (2010) 43–48.