Ego Networks in Twitter: an Experimental Analysis

Conference Paper *in* Proceedings - IEEE INFOCOM · April 2013

DOI: 10.1109/INFCOM.2013.6567181

4 authors:

Valerio Arnaboldi
California Institute of Technology

Marco Conti
Italian National Research Council

Andrea Passarella
Italian National Research Council

Fabio Pezzoni
Italian National Research Council

Ego Networks in Twitter: an Experimental Analysis

Valerio Arnaboldi, Marco Conti, Andrea Passarella and Fabio Pezzoni
Institute of Informatics and Telematics of CNR
via G. Moruzzi, 1 - 56124 Pisa, Italy
Email: {v.arnaboldi, m.conti, a.passarella, f.pezzoni}@iit.cnr.it

***Abstract*—Online Social Networks are amongst the most important platforms for maintaining social relationships online, supporting content generation and exchange between users. They are therefore natural candidate to be the basis of future human-centric networks and data exchange systems, in addition to novel forms of Internet services exploiting the properties of human social relationships. Understanding the structural properties of OSN and how they are influenced by human behaviour is thus fundamental to design such human-centred systems. In this paper we analyse a real Twitter data set to investigate whether well known structures of human social networks identified in "offline" environments can also be identified in the social networks maintained by users on Twitter. According to the well known model proposed by Dunbar, offline social networks are formed of circles of relationships having different social characteristics (e.g., intimacy, contact frequency and size). These circles can be directly ascribed to cognitive constraints of human brain, that impose limits on the number of social relationships maintainable at different levels of emotional closeness. Our results indicate that a similar structure can also be found in the Twitter users' social networks. This suggests that the structure of social networks also in online environments are controlled by the same cognitive properties of human brain that operate offline.**

I. Introduction

The last generation of electronic devices (e.g., smartphones, tablets) offer plenty of different ways to generate and share information inside the electronic world, formed of virtual contacts and relationships. Online Social Networks (OSN - e.g., Twitter, Facebook and Google+) represent the most advanced means of communication at our disposal to socialise within virtual environments. In addition, they already provide one of the most used platform that enable users to share their own generated content in their social communities. Content sharing is only one example of an Internet service that is tremendously benefiting from information about users' social relationships, available through OSN. In perspective, OSN can thus be seen as natural platforms to support novel Internet networking and communication services (well beyond content sharing) that are based on, and exploit social relationships maintained by users in online environments. Broadly speaking, as a preliminary example, it has been already proposed to build P2P unstructured networks based on common interests and social relationships between users [1], [2]. We think that this human-centred approach has a great potential, and can be exploited more extensively in the design of networks and services for the Future Internet. In fact it allows network designers to take advantage of the behaviour of the most important users of the network, i.e. humans. To be able to design such type of services and engineer the networking environments that will support them, achieving robustness, efficiency and effectiveness in the communication, we must understand how people relate in OSN and what are the structural properties of social relationships in OSN.

A very intriguing question is whether social networks maintained in OSN are structurally different from social networks maintained by humans in "offline" environments, i.e. by using communication means that do not involve OSN services (such as face-to-face contacts, mail exchange, phone calls, ...). As R.I.M. Dunbar - a British anthropologist - pointed out, offline social networks properties and structures are directly shaped by the the cognitive constraints of human brain [3]. Even though OSN have been extensively analysed hitherto under many different aspects, a complete understanding of the structural properties of the users' social networks is still to be achieved. In particular, it is unclear if OSN structures differ from those found by Dunbar and other scientist in offline environments. On the one hand, a similarity between online and offline social structures could indicate that the use of OSN is not changing the capability of our brain to maintain social relationships. On the other hand, since communication means offered by OSN are so different than those used in offline environments, OSN could have completely different structures from those found offline.

The aim of this paper is to investigate whether it is possible to find evidences of the presence of social structures in OSN. Specifically, we analyse a Twitter data set we have collected, containing communication records of more than $300,000$ accounts, seeking whether it is possible to identify social structures similar to those discovered in offline networks. We also make a detailed classification of our data set to identify people using Twitter mainly to communicate and maintain their social relationships, separating them from other kinds of accounts (companies, news brokers, ...) not relevant for our study. This classification turned out to be essential for a correct understanding of the social dynamics behind virtual social relationships. Our results indicate that social structures quite similar to those found in offline environments can also be identified in Twitter. In particular, we find groups of social relationships at different levels of intimacy, with quantitative properties similar to those found offline. This confirms similar results we have previously obtained by analysing a large Facebook data set [4]. The methodology used in the present paper is similar to that used in [4]. Nevertheless, the scientific contribution of the present work goes far beyond, since the similarities found between Twitter and Facebook confirm that our results are valid for OSN in general and are not limited to a particular platform. Twitter and Facebook have many structural differences, particularly for what concerns the kind of supported communication and the management of social relationships. Although this, it is remarkable that the typical social pattern found offline is clearly identified in both OSN. This suggests that, even if OSN significantly differ from more traditional communication means, the constraints of our brain impose the presence of the same social structures and the same limits we encounter offline. From our analysis, we are also able to isolate a small group of relationships that has not yet been included in the taxonomy introduced by Dunbar, but that seems to be a natural constant in human species. These relationships are the links that people maintain with their most important contacts (one or two on average), perhaps a partner and/or a best friend. This is also an interesting result. Under the assumption that structural properties of social networks do not change across offline and online environment, this shows one of the advantages of studying social networks in OSN platforms, which typically provide much larger data sets with precise characterisation of the features of interactions between users (and thus, in principle, more precise social information) with respect to studies done using offline interaction data.

The paper is organised as follows: in Section II we introduce the basic concepts used to define social network structures in anthropology and sociology and we describe what has been done hitherto to analyse the structure of OSN. Then, in Section III, we describe the data set we have used in the analysis. In Section IV we describe the classifier we used to filter the data set, selecting the Twitter accounts useful for our study. Section V introduces the methodology used to characterise the social structures in Twitter and to compare them with offline networks and other OSN. In this section we also present our results, discussing their meaning and comparing them with other background findings in different environments. Finally, in Section VI we draw the main conclusions of this analysis.

II. Background and Related Work

Before the widespread of ICT, properties and dynamics of human social networks were studied by social anthropologists using logs of communication means such as mail exchanges, post cards, phone calls, etc. One of the most important result in this field was the discovery of a strong relation between the amount of cognitive resources and the number of social relationships an individual can actively manage. Basing on the correlation between the size of the neocortex (a part of the cerebral cortex) in primates and the size of their social groups, anthropologists hypothesised that the human brain provides cognitive resources for maintaining about 150 active social relationships [5]. This limit is popularly known as the *Dunbar's number* and its presence was validated by independent social experiments [3].

Another important consequence of the existence of cognitive constraints is the hierarchical organisation of social relationships. In fact, studies demonstrated that people tend to maintain social relationships at different levels of intimacy.

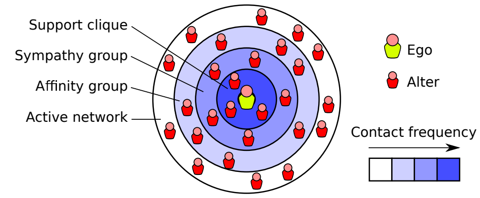

Fig. 1. Ego network model.

Dunbar identified four main classes of social relationships, each of them characterised by a typical level of intimacy. These classes can be represented by simple social network model called *ego network*. An ego network consists of an individual (the *ego*) and all the persons the ego has a social relationship with (the *alters*). In this representation, the classes of relationships discovered by Dunbar can be thought as inclusive *circles*, as depicted in Fig. 1. The inner circle, called *support clique*, represents the strongest social relationships of the ego. Outer circles, characterised by increasing size and decreasing level of intimacy, are called *sympathy group*, *affinity group* and *active network*. Although the size of the circles can vary significantly across individuals (and different methodologies used to collect data), it is possible to identify their average sizes as 5, 15, 50 and 150 respectively, featuring a characteristic scaling factor approximately equal to three [3], [6].

While the size of the circles discovered by Dunbar (hereafter referred as *Dunbar's circles*) are relatively easy to measure, the related levels of intimacy are not directly definable due to their abstract nature. To overcome this limitation, researchers commonly characterise the Dunbar's circle in terms of minimum frequency of contact between the ego and the alters. The support clique is thus defined as the people the ego contacts at least weekly, the sympathy group as the people contacted at least monthly and the active network as the people contacted at least yearly. The affinity group is not clearly defined in literature in terms of minimum frequency of contact. Note that a larger number of social acquaintances can also be identified outside the active network, characterised by a smaller contact frequency. These are relationships that, by being not sustained by sufficient interaction frequency, do not consume cognitive constraints, but whose social significance is definitely minimal [6].

The proliferation of OSN led researchers to investigate if the social behaviour of people in online environment is comparable with observations obtained in offline environments. For instance, evidences of the presence of the Dunbar's number in online environment were discussed in two different studies. Specifically, in [7] a small sample of Facebook users participated to a survey in which they were invited to rank their online friendships. Analysing the survey data, authors conclude that the number of Facebook friendships users consider as relevant is, on average, 108.17 (which is of the same order of the Dunbar's number). In [8], authors analysed a large data set extracted from Twitter and discovered that users with a number of active online friendships between

100 and 200 maximise the average number of replies sent per friend. Beyond this threshold the cognitive constraints limit the possibility to maintain relationships at the same level of intimacy. This was considered as an indication of a social capacity constraint at a network size similar to the Dunbar's number. According to the discussion in [9], the number of replies sent or received by users is a good index to identify "real" friends and to discard meaningless links.

Another important study on the similarities between offline and online social networks is presented in [4]. Authors consider a large Facebook data set from which it is possible to obtain the frequency of the interactions between pairs of users. After the selection of a significant subset of user accounts, authors apply standard clustering techniques on the interaction frequency distribution of each user obtaining the underlying structure of their ego networks. Comparing this structure with the model described by Dunbar, similarities emerge in the number of circles (four), their dimensions, and the associated scaling factors.

In this work we present a study aimed at characterising structures of ego networks in Twitter. Since Twitter encompasses different categories of users, in Section IV we isolate the users which are relevant for our analysis, building a classifier. Other work in literature has been done to classify different kinds of Twitter users. For instance, in [10], a classifier is built in order to distinguish between "humans", "bots" and "cyborgs". The classifier takes into account the entropy level of the communication, text patterns in tweets and account-related variables. The evaluation of the system, using a manually-classified test set, shows that automated classifier can achieve a valuable level of accuracy in Twitter. Our classifier, differently than the one in [10], focuses on the distinction between "socially active people" and other kind of accounts. This distinction, essential in our analysis, is performed using a reduced number of variables, without the need to access the textual content of the tweets.

III. Dataset Description

We implemented a crawling agent which is able to download user profiles and their communication data from Twitter. The agent visits the Twitter graph considering the users as nodes and their contacts as links. In particular, we assume that a link between two nodes exists if at least one of the users follows the other or an interaction between them has occurred. We use as indications of interactions the presence of *mentions* in a tweet (i.e., the fact that a user explicitly mentions the other in a tweet) and *replies* (response to previous mentions). In other words, we consider only relationships showing a certain level of bi-directionality. Social relationships are in general asymmetric (i.e. the importance of a social link can be different for the two users involved in that link), but, in any meaningful social relationship, some level of bi-directionality must be present.

The crawling agent starts from a given user profile (seed) and visits the Twitter graph following the links. For each visited node, we take advantage of the Twitter REST API to extract the user *timeline* (i.e., the list of tweets she posted including mentions and replies), her *following* (i.e., people she follows) and her *followers* (i.e., people who follow her). Twitter REST API limits the amount of tweets that can be downloaded per user to $3,200$. This does not represent a constraint to our analysis since, as we show in the following, it is sufficient for our purposes.

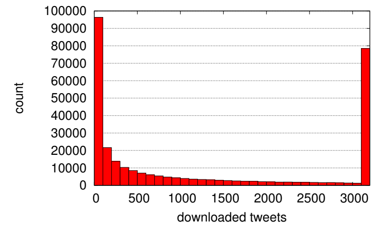

Fig. 2. Downloaded tweets per user distribution.

The crawling agent uses 250 threads that concurrently access a single queue containing the ids of the user profiles to download. Each thread extracts a certain number of user ids from the queue, then it gets the related profiles and communication data from Twitter using the REST API. Finally, after extracting new user ids from the communication data and from the following/follower lists, the threads add them to the queue. The use of multiple threads allows both to speed-up the data collection and to avoid the crawler to remain trapped in visiting the neighbourhood of a node with a large number of links. The seed we used to start the data collection is the profile of a widely know user (user ID: 813286), so that her followers represent an almost random sample of the network.

We have collected a data set from $303,902$ Twitter users, whose data was downloaded in November 2012. In the column "all" of Table I we present some statistics of the data set, while in Fig. 2 we show the distribution of the number of tweets downloaded per user. In the figure we can notice the presence of a peak in correspondence of the value $3,200$ that is the maximum amount of tweets downloadable using the Twitter REST API. Cases where the number of tweets is lower than $3,200$ correspond to users that have generated less than $3,200$ tweets since their account has been created. The number of users that posted an amount of tweets above this threshold is indicated in the table by $N_{3,200}$. In the table we also report the average number of tweets, following and followers per user and the average ratio of replies and tweets containing mentions (over the total number of tweets). Each average value is reported with 95% CI between square brackets. Data reported in the table indicate that around 20% of the tweets downloaded by our crawler contain direct communication between people, important for our study. This percentage is sufficient to perform significant analysis on the data set.

IV. Classification

Differently from other online social networking services, Twitter is designed to encompass heterogeneous types of users. In fact, in addition to accounts used by persons mainly to communicate and maintain their social relationships with others (hereafter referred to as *socially relevant people*), there exist Twitter accounts representing companies, public figures, news broadcasters, bloggers and many others, including spammers and bots.

Since we are interested in characterising social aspects of human relationships in online environments, we implemented an automatic procedure to distinguish between socially relevant people from all the other accounts. To this aim, we decided to make use of a supervised classifier dividing Twitter accounts in two classes labelled “people” and “others” respectively. We manually classified a sample of 500 accounts, randomly drawn from the data set, and we used this classifications to train a Support Vector Machine [11]. This SVM uses a set of 115 variables: 15 of them related to the user’s profile (e.g., number of tweets, number of following and followers, account lifespan) and 100 obtained from her timeline (e.g., percentage of mentions, replies and retweets, average tweets length, number of tweets made using external applications).

To test the generality of the SVM (i.e., the ability to categorise correctly new examples that differ from those used for training) we take 10 random sub-samples of the training set, each of which contains 80% of the entries, keeping the remaining 20% for testing. Then, we apply the same methodology used to create the SVM generated from the entire training set on the 10 sub-samples. Doing so, we obtain different SVMs, trained using different sub-samples of the training set, and of which we are able to assess the accuracy. The average accuracy of these SVMs can be seen as an estimate of the accuracy of the SVM derived from the complete training set. Specifically, we calculate the *accuracy* index, defined as the rate of correct classifications, and the *false positives rate*, where false positives are accounts wrongly assigned to the “people” class. In our analysis we consider only users falling in the “people” class, thus it is particularly important to minimise the false positive rate[1]. Minimising the false negative rate is also important but less critical, as false negatives result in a reduction of the number of users on which we base our analysis.

The average accuracy of our classification system is equal to 0.813 $[0.024]$ and the average false positives rate is 0.083 $[0.012]$ (values between brackets are the 95% CI). These results indicate that we are able to identify socially relevant people in Twitter with sufficient accuracy, even if people have different behaviours and characteristics (e.g., different culture, religion, age). Moreover, the false positive rate is quite low (below 10%). The results are of the same magnitude as those found in a similar classification performed in Twitter [10].

After applying the classifier to the whole data set we have extracted $205,108$ socially relevant people. Some properties of the classes “people” and “others” are reported in Table I. It is worth noting that users in the class “others”, on average, have a much higher number of following and followers compared to the users in the class “people”. Similarly, we have downloaded

[1]False negatives are probably “people” with behaviour similar to the subjects in the “others” class. For this reason we consider them as outliers, since our analysis is focused on mean values of social properties of the users

TABLE I
DATA SET (ALL) AND CLASSES STATISTICS

	all	**people**	**others**
N	$303,902$	$205,108$	$98,794$
$N_{3,200}$	$77,196$	$38,107$	$39,088$
($\%$ $N_{3,200}$)	(25.4%)	(18.6%)	(39.6%)
$\#$ $tweets$	$1,234$ $[5]$	979 $[5]$	$1,764$ $[8]$
$\#$ $following$	$1,905$ $[33]$	673 $[8]$	$4,462$ $[98]$
$\#$ $followers$	$11,335$ $[529]$	777 $[107]$	$33,254$ $[1,602]$
$\%$ $tweets_{\mathrm{REPL}}$	17.4%	18.4%	15.4%
$\%$ $tweets_{\mathrm{MENT}}$	22.7%	21.6%	24.7%

a higher number of tweets from the user belonging to “others” than from the users in “people”. In the table we also report the comparison of other important variables, extracted from the classes, which exhibit significant dissimilarities. While the users in “people” class have a higher use rate of replies, user in “others” show a higher usage of mentions. Even though the number of tweets we downloaded for the two classes do not significantly differ, the number of accounts with more than $3,200$ tweets is much higher in “others” class. These results are aligned with the intuition about the different use of Twitter by humans to maintain social relationships, with respect to other type of users, in particular commercial and political ones. Specifically, “people” tweet less than “others” and have (far) less following and followers. It is also interesting to note the higher percentage of replies, which is an indication of a more marked attitude towards bidirectional interactions, which is also an intuitive difference between the two classes.

V. ANALYSIS OF TWITTER EGO NETWORKS

lWe use the “people” data set to analyse the structure of ego networks in Twitter. We consider each Twitter user in the data set as an ego, and we define as *friend* a person in Twitter to whom the ego has sent at least one reply. A reply implies bi-directional communication and indicates that both the user and her friend has spent a certain amount of their cognitive and time resources to interact.

A. Users’ Interaction as a Function of Ego Network Size

The first analysis we make is to study the average number of replies sent by the users to their friends. Specifically, in [8], this was the main index used to conclude that a concept similar to that of the Dunbar’s number (the maximum number of active social relationships a human can maintain) holds also in Twitter ego networks. Analysing this index allows us to understand whether our data set is aligned with the one used in [8] as far as this index is concerned.

Fig. 3 depicts the trend of the average number of replies per friend as a function of the number of friends of the user. Differently from [8], we divide the analysis for the two classes identified in Section IV: “people” and “others”. The results, supported by the figure, highlight a clear distinction between the properties of the two classes.

The class “people” shows a higher mean value of replies per friend and a maximum around 80 friends. This is an indication of the effect of the cognitive limits of human

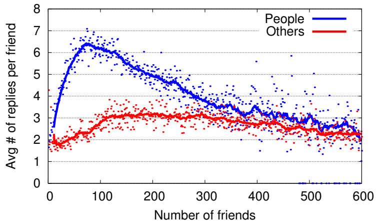

Fig. 3. Average number of replies made by accounts with different number of friends

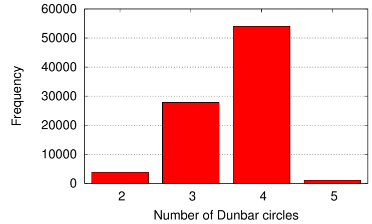

Fig. 4. Distribution of the number of Dunbar's circles

brain on the ability to maintain social relationships in online social networks. The peak of the curve identifies the threshold beyond which the effort dedicated to each social relationship decreases. This is due to the exhaustion of the available cognitive/time resources that, therefore, have to be split over an increasing number of friends. As discussed in [8], this can be seen as an evidence of the presence of the so called Dunbar's number in Twitter.

The class "others" shows a more random pattern, with lower average value of replies per friend without any significant discontinuities. This indicates that the accounts belonging to the class "others" are not influenced by cognitive capabilities. In fact they are often managed by more than one person or by non-human agents.

B. Structure of Ego Networks: Methodology

After this analysis we perform a refined selection of the ego networks in the data set to identify the most relevant set of accounts for our study. Specifically, we used a methodology similar to the one used in [4] to analyse ego networks in Facebook. We discard too recent accounts (i.e., created less than 6 months before the time of the download) since we think they are are not long enough to allow users to create a meaningful ego network, i.e. to select friends in the ego network and communicate with them long enough to well reflect the level of intimacy of the relationship. For the same reason, we discard friendships with a duration lower than one month.

After the selection of the ego networks relevant for the analysis, we focus on identifying their structure. To this aim we use standard clustering techniques to find out if social relationships with friends could be grouped according to their frequency of contact. We obtain these frequencies by dividing the number of replies sent to each friend by the duration of the considered friendship (i.e., the time since the first mention or reply sent to the friend). Hence, a quantitative analysis of the properties of the groups of relationships would be essential to highlight analogies and differences with offline ego networks structure and the results found in Facebook [4]. To this aim we use the $k-$means algorithm. With the $k-$means algorithm we partition the frequencies of contact of each ego network into a fixed number (k) of different clusters, according to their Euclidean distance. To find the number of clusters in each ego network we repeatedly apply $k-$means with increasing values of k. For each value of k, the standard k-means technique provides an index between 0 and 1 that measures the quality of the obtained clustering. This index monotonically increases with k. It is a standard technique to assume as optimal k the one beyond which increasing k yields an increase of the index below a given threshold. We set this threshold to 0.1, as done in [4], to be able to obtain comparable results.

C. Structure of Ego Networks: Analysis

From the clusters obtained applying $k-$means on the frequencies of contact of each ego network we obtain the inclusive structures known in literature as Dunbar's circles [3]. Namely, each circle is represented by joining the respective cluster and all its sub-clusters (i.e., the clusters with higher contact frequency).

The distribution of the characteristic number of Dunbar's circles of ego networks in the data set, depicted in Fig. 4, shows that most of the ego networks have 4 Dunbar's circles. Specifically, the average number of circles is equal to 3.14 and its median is 4. Moreover, in Table II we report some statistics (with 95% CI between square brackets) about ego networks aggregated for different number of circles we have found. It is worth noting that, as the number of circles increases, the average network size and the average Twitter use rate (defined as the average frequency of contact multiplied by the number of friends) also increases. The Twitter use rate is a proxy for the amount of time a user spends in Twitter, that is to say the budget of time the user allocates for socialising in Twitter. Another interesting finding is that the ego networks with 4 circles are those with the highest average number of replies sent per friend. According to the methodology used in [8], this marks the point where the cognitive "capacity" allocated to social relationship is saturated.

According to the results, for most of the ego networks in Twitter we identify the same number of Dunbar's circles (four) found in offline ego networks and in other online environments [4]. Nevertheless, it is noteworthy that there is a non negligible amount of people in Twitter with only three circles. This confirms what we have previously found in Facebook [4] and it is a strong indication of the presence

TABLE II
PROPERTIES OF EGO NETWORKS WITH DIFFERENT NUMBER OF CIRCLES

Circles	# egos	Net size	Use rate	# replies
2	$3,819$	3.04 [0.01]	2.86 [0.29]	2.71
3	$27,788$	38.05 [0.83]	62.63 [1.48]	5.19
4	$53,982$	80.31 [0.86]	113.28 [1.33]	5.35
5	$1,073$	190.03 [14.31]	167.06 [12.15]	3.81

of two different kinds of users in online social networks: (i) *occasional users*, with a small three-circles ego network and a low Twitter use rate, and (ii) *active users*, with an higher use rate and a number of Dunbar's circles similar to that found in offline environments. This distinction is similar to the difference between more- and less-sociable individuals found in offline environment [12].

We further analyse the properties of the identified Dunbar's circles in Table III, where we apply k-means, with $k = 4$, to all the ego networks in the data set. The contact frequency is measured in number of replies per month. The typical contact frequencies of the circles (i.e., the minimum frequencies needed for a relationship to be part of the circles) are $\sim$ once every two days for C_1, $\sim$ *weekly* for C_2, $\sim$ *monthly* for C_3 and $\sim$ *twice a year* for C_4. It is remarkable that some of the circles found in Twitter show properties similar to those found in offline social networks. In particular, C_2 and C_3 respectively resemble the support clique and the sympathy group in terms of size and contact frequency. C_4, according to its properties, can be placed between the affinity group and the active network. These results are compatible with the models for offline ego networks [3], [6] and Facebook ego networks [4] also as far as the scaling factor between the circles, which is approximately equal to 3.

From the analysis we have performed, an additional circle (C_1) emerges. It is typically formed of one or two people strongly connected to the ego. This circle could be seen as a *super support clique*, and the friends contained in it are the most important relationships for the ego, perhaps a partner and/or a best friend. Scientists have long predicted the existence of this circle but they have never been able to prove it[2], due to the limitation of the methods used in offline analysis. The existence of such an additional circle, although needs to be supported by more detailed analyses, provides a very interesting result from the standpoint of the study of human social networks, and show a concrete example of the potential of characterising them through data collected on social networking sites.

Our results also indicate that the size of the last Dunbar's circle (called active network) is smaller than the reference value found in offline environments. Looking carefully at it, its contact frequency appears to be lower than the affinity group ($\sim$ eight times a year [4]), but higher than the active network described by Dunbar (once a year). The small size of this circle could be conditioned by the use of Twitter replies to weight social relationships. This index may not be the best choice to measure weak relationships, as it emphasises a lot bi-directionality of interaction, which may be less present in weak relationships (than in strong ones). Nevertheless the size of this circle is compatible with other results in literature about offline networks [13].

VI. Conclusions

In this paper we analysed a real Twitter data set containing a high number of communication records to investigate the structure of ego networks in Twitter. We separated "people" between other kinds of accounts (called "others") to effectively study the properties of social relationships in Twitter. We applied standard clustering techniques to the frequencies of contact extracted from the data set to characterise the Dunbar's circles in Twitter. The results indicate that Twitter presents social structures qualitatively similar to that found by Dunbar in offline ego networks and by ourselves in a similar study on Fcebook [4]. This suggests that Twitter (as we have previously shown for Facebook) does not fundamentally change the structure of human ego networks, which is instead determined by other characteristics of the human socialising process, such as the maximum amount of cognitive resources dedicated to social activities [6]. Additionally, we have found a very small circle, not present in the taxonomy introduced by Dunbar, formed of, on average, one or two people with an extremely strong social relationship with the ego. Moreover, the active network size in Twitter seems smaller than that found offline. These results indicate, on the one hand, a strong similarity between online and offline social networks and, on the other hand, the presence of additional properties not visible in offline networks. Even though these new properties have not yet been investigated in sociology, they have an intuitive meaning in humans and they should be further investigated to understand their role in social networks.

[2]As stated by R.I.M. Dunbar in a private communication on June 19, 2012.

TABLE III
PROPERTIES OF 4-CIRCLES TWITTER EGO NETWORKS

	$\mathbf{C_1}$		$\mathbf{C_2}$		$\mathbf{C_3}$		$\mathbf{C_4}$
Size	1.74 [.03]		5.75 [.07]		17.56 [.21]		70.04 [.69]
Scaling fc.		3.31		3.06		3.99	
Contact fr.	17.28		6.00		1.77		0.20

Acknowledgment

This work was partially funded by the European Commission under the SCAMPI (FP7-FIRE 258414), RECOGNITION (FP7 FET-AWARENESS 257756), and EINS (FP7-FIRE 288021) projects.

References

[1] M. Conti and M. Kumar, "Opportunities in opportunistic computing," *Computer*, vol. 43, no. 1, pp. 42–50, 2010.

[2] V. Arnaboldi, M. Conti, and F. Delmastro, "Implementation of CAMEO: a Context-Aware MiddlEware for Opportunistic Mobile Social Networks," in *12th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WOWMOM)*, 2011.

[3] S. G. Roberts, "Constraints on Social Networks," in *Social Brain, Distributed Mind (Proceedings of the British Academy)*, 2010, pp. 115–134.

[4] V. Arnaboldi, M. Conti, A. Passarella, and F. Pezzoni, "Analysis of Ego Network Structure in Online Social Networks," in *ASE-IEEE International Conference on Social Computing (SocialCom)*, 2012.

[5] R. I. M. Dunbar, "The social brain hypothesis," *Evolutionary Anthropology*, vol. 6, no. 5, pp. 178–190, 1998.

[6] A. Sutcliffe, R. Dunbar, J. Binder, and H. Arrow, "Relationships and the social brain: Integrating psychological and evolutionary perspectives." *British journal of psychology*, vol. 103, no. 2, pp. 149–68, 2012.

[7] V. Arnaboldi, A. Passarella, M. Tesconi, and D. Gazzè, "Towards a Characterization of Egocentric Networks in Online Social Networks." in *OTM Workshops*, vol. 7046, 2011, pp. 524–533.

[8] B. Gonçalves, N. Perra1, and A. Vespignani, "Modeling Users' Activity on Twitter Networks: Validation of Dunbar's Number," *PLoS ONE*, vol. 6, no. 8, Aug. 2011.
[9] E.-A. Baatarjav, A. Amin, R. Dantu, and N. K. Gupta, "Are You My Friend?" *7th IEEE Consumer Communications and Networking Conference (CCNC)*, pp. 1–5, Jan. 2010.
[10] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Who is tweeting on twitter: human, bot, or cyborg?" in *Annual Computer Security Applications Conference*, 2010.
[11] C. Cortes and V. Vapnik, "Support-vector networks," in *Machine learning*, 1995, pp. 273–297.
[12] R. I. M. Dunbar and M. Spoors, "Social Networks, Support Cliques and Kinship," *Human Nature*, vol. 6, no. 3, pp. 273–290, 1995.
[13] S. G. Roberts, R. I. Dunbar, T. V. Pollet, and T. Kuppens, "Exploring variation in active network size: Constraints and ego characteristics," *Social Networks*, vol. 31, no. 2, pp. 138–146, 2009.