Dynamics of Personal Social Relationships in Online Social Networks: a Study on Twitter

Valerio Arnaboldi, Marco Conti, Andrea Passarella
IIT-CNR, via G. Moruzzi, 1 - 56124 Pisa, Italy
{v.arnaboldi, m.conti, a.passarella}@iit.cnr.it

Robin Dunbar
Department of Experimental Psychology, University of Oxford
South Parks Road, Oxford OX1 3UD, United Kingdom
robin.dunbar@psy.ox.ac.uk

ABSTRACT

The growing popularity of Online Social Networks (OSN) is generating a large amount of communication records that can be easily accessed and analysed to study human social behaviour. This represents a unique opportunity to understand properties of social networks that were impossible to assess in the past. Although analyses on OSN conducted hitherto revealed some important global properties of the networks, there is still a lack of understanding of the mechanisms underpinning these properties, their relation to human behaviour, and their dynamic evolution over time. These aspects are clearly important to understand and characterise OSN and to identify the evolutionary strategy that favoured the diffusion of the use of online communications in our society.

In this paper we analyse a data set of Twitter communication records, studying the dynamic processes that govern the maintenance of online social relationships. The results reveal that people in Twitter have highly dynamic social networks, with a large percentage of weak ties and high turnover. This suggests that this behaviour can be the product of an evolutionary strategy aimed at coping with the extremely challenging conditions imposed by our society, where dynamism seems to be the key to success.

Categories and Subject Descriptors

H.3.4 [**Information Storage and Retrieval**]: Systems and Software—*information networks*; H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*web-based services*

General Terms

Measurement, Human Factors

COSN'13, October 7–8, 2013, Boston, Massachusetts, USA.

http://dx.doi.org/10.1145/2512938.2512949.

Keywords

Online social networks, Ego networks, Personal social relationships

1. INTRODUCTION

Online Social Networks (hereafter OSN) are one of the most important communication means that we use in our everyday life. They help us to maintain our social relationships with family and friends, as well as to enlarge our professional sphere and to acquire knowledge and new ideas from the network. OSN popularity is due to their ability to transform people into *active* producers of information, letting them create, access and share contents anywhere and anytime.

These unique characteristics of OSN are producing strong effects on our society, but the extent to which they are impacting on human social behaviour is still unknown. Nevertheless, there is no doubt that their role will be of primary importance in our future. For this reason, studying people's behaviour in OSN is of great value to understand how the society is evolving and how we can contribute to the process, designing future OSN able to fulfil users' needs in terms of management of social relationships through digital communications.

In this paper we analyse a Twitter data set containing communication traces of more than 2 million users to study the dynamic properties of the behaviour of OSN users and to start indicating analogies and differences between *online* and *offline* social networks, comparing our results with the findings in literature about more traditional types of social networks (e.g., face-to-face or phone calls social networks).

The novelty of our work, compared to other analysis performed on OSN, resides in the uniqueness of our data set, and in the focus on the dynamic evolution of social structures over time. In fact, we were able to obtain the last 3,200 tweets of a large data set of users, that is in most cases enough to represent their whole communication history. This allowed us to carry out a sensitive analysis about the evolution of human social behaviour in Twitter over time (our data set covers user activities over a time span up to 7 years, and, on average, of one year). This new approach to studying the *dynamic* properties of social relationships and networks revealed many important aspects of OSN that should be considered to correctly understand their social properties. To the best of our knowledge, this is the first work that provides an extensive characterisation of the dynamic evolution of human social structures in one of the reference OSN.

The analysis of the evolution of human social behaviour in OSN has several practical implications. For example, it could be the basis of innovative applications that dynamically track the structure of the social networks of the users, helping people in the maintenance of their social relationships and suggesting possible actions to improve their social experience. Or, it could be used to classify users based on their dynamic behaviour (we show that different classes of users can be identified), and use this classification as context information for customising other OSN applications. In general, it can be used for personalising the OSN applications experience to the specific dynamic social behaviour of the users.

The results of our analysis indicate that in Twitter people behave in a significantly different way than in other kinds of social networks. One of the key results we present is that people prefer to maintain weak social relationships than strong ones, with a high turnover of contacts in their networks. This behaviour fits perfectly in our extremely dynamic society, where people must quickly adapt to cope with frequent changes in their life, from their sentimental sphere to their work. OSN like Twitter seem thus to be useful tools to have more access to new resources from the network and to manage light-weight social relationships, easy to be created, maintained and destroyed when needed. For this reason we think that this use of OSN can be seen as part of an evolutionary strategy adopted by humans to cope with the very dynamic conditions of the society we live in. Note that these results become evident only when the *dynamic* evolution over time of OSN social relationships is studied, while they remain "hidden" in static analyses (like those available in the literature) that observe the aggregate properties of social relationships over long time intervals. In addition, we also highlight the existence of different types of users, that can be broadly divided into two main categories: people who have a short, but intense, activity and people who interact with social peers for long time intervals. Different properties can be highlighted for these different classes, with the latter having smaller, but more stable, networks than the former, and much more similar to social networks found in previous analyses (such as [31]).

Finally, our results also suggest that, while there is a large number of users that abandon Twitter after a relatively short amount of time, there is a significant fraction of users that keep using it mostly at a constant rate. This suggests that the hypothesised decline in the use of OSN [26, 29] may not be present (at least in Twitter).

The paper is organised as follows: in Section 2 we introduce the related work in literature about the study of human social behaviour in social networks, from the point of view of different disciplines that analyse the subject. In Section 3 we describe the data set we collected and analysed. Then, in Section 4 we describe the methodology we used to study the data set. Hence, in Section 5 we present the main results of the analysis. Finally, in Section 6 we draw the conclusions of our work.

2. RELATED WORK

In this Section we summarise the main results about the characterisation of human behaviour in social networks found in different research fields. We classify social networks into two different categories: *offline* and *online* social networks. With the term "offline" we refer to all the social networks maintained with traditional (i.e., non-digital) communications. On the other hand, "online" is referred to social networks maintained by using digital communications (e.g., e-mail, social media applications, phone calls). This distinction helps us to identify the difference between the two worlds, and to understand how the introduction of digital communications shaped human social behaviour.

2.1 Offline social networks

The study of social networks started from the analysis of offline networks, typically extracted from questionnaires data and interviews. These kind of social networks have been primarily analysed in sociology, anthropology and evolutionary psychology.

2.1.1 Social network analysis

Social networks analysis (SNA) emerged from sociology in the 20th century. The first pioneers of SNA defined social networks as an ensemble of ties denoting the existence of a social connection between two individuals. From sociology, many important aspects of social networks have been found. Mark Granovetter discovered that our social contacts have different characteristic properties and their strength is unevenly distributed in the network. Strong ties are maintained with people close to us, while weak ties usually represent bridges between different communities and are thus important for accessing new ideas and resources [15]. Granovetter, in his seminal work, gave also an informal definition of tie strength, that is still used in many different analyses. Peter Marsden was one of the first to test the definition of tie strength given by Granovetter applying an analytic model on real data [20]. His findings revealed a strong correlation between the terms "tie strength" and "emotional closeness". Ronald Burt discovered that the social capital - that can be assimilated to the concept of quantity of resources acquired from the network - is negatively influenced by the presence of discontinuities in the distribution of social links in the network, called structural holes [5]. If a person can broker connections between otherwise disconnected segments her social capital increases.

Another important property of social networks is the average *distance*, the average shortest path length between any tho people in the network. Stanley Milgram, in his famous experiment, found that the typical distance in a social network is around 6. This property is better known as the "small world" or the "six degrees of separation" [27].

In general, sociologists focused their attention on the structural properties of the network, explaining the relation between these properties and human behaviour. For a complete description of all the known properties of social networks seen from a technical point of view we refer the reader to [10].

2.1.2 Ego networks and evolutionary psychology

A different approach to the study of social networks has been adopted by anthropologists and evolutionary psychologists. Rather then focusing on the global properties of the network they look at the local properties of personal social networks, often called *ego networks*. An ego network (depicted in Figure 1) is a simple model that describes the

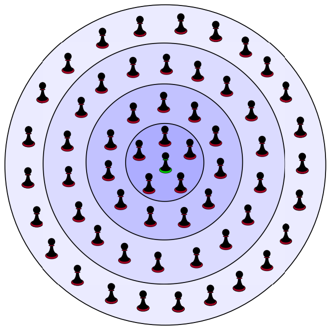

Figure 1: Ego network model

social relationships between an individual, called *ego*, and all the contacts ego has with other people, called *alters* - as defined by the standard notation in ego network analysis. The tie strength between ego and alters is modelled as the distance between them and is often estimated using the frequency of contact or the time since last contact between the people involved in the relationship [4, 16].

The most important result found on ego networks is that there is a limit on the number of alters people can actively maintain in their network, due to the cognitive constraint of human brains. This limit has been discovered by Robin Dunbar, who pioneered the study on primates, finding a positive correlation between the neocortex size - the part of the brain responsible for social activity - of different type of apes and the size of their social groups [9]. Dunbar predicted that the size of social groups in humans, given their large brain size, should be around 150 - this number is known as the *Dunbar's number*. Dunbar's hypothesis has been validated with many different experiments in offline and online social networks [4, 13, 31].

Another important result found on ego networks is that alters around ego form a layered structure with a series of inclusive concentric circles with typical characteristics and size (depicted in Figure 1). Ego can be envisioned as sitting in the centre of these circles and alters are placed around her depending on their emotional closeness. Inner circles have higher emotional closeness and frequency of contact, but lower size, since maintaining these strong relationships is extremely expensive. Moving from the inner to the outer layers, the emotional closeness decreases and the size of the layers increases, due to the lower cost of the relationships.

In anthropology and psychology literature four layers are usually identified in ego networks, as depicted in Figure 1. The first layer, called *support clique*, contains the alters to whom ego seeks advice in case of strong emotional distress or financial crisis and is usually limited to an average of 5 members. The other layers are called *sympathy* group (15 members), *affinity group* (50 members) and *active network* (150 members). The active network corresponds to the set of people that are actively maintained by ego in her network, identified by the Dunbar's number. The remarkable property of these layers is that the scaling ratio between the size of adjacent circles appears to be a constant in humans and is close to 3 [31]. The representative frequency of contact associated with these layers is *one message a week* for the support clique, *one message a month* for the sympathy group and *one message a year* for the active network. The properties of the affinity group are not accurately defined in anthropology and have been only recently investigated in OSN [2, 3]. Other layers exist beyond the limit of the Dunbar's number, but the social relationships they contain represent only acquaintances, for which ego does not actively invest cognitive resources. For a complete ethnographic definition of the different ego network layers see [25].

Ego networks in evolutionary psychology have been studied mainly to understand the cause that induced humans to develop a large and extremely expensive brain compared to other animals. The analysis of social networks lead to the "social brain hypothesis" (SBH). SBH identifies the key factor at the basis of the evolution of human brain as the growing need for our ancestors to maintain an increasing number of social relationships with different groups to survive in the extremely challenging environmental conditions arose during the last ice age [9].

During their evolution, animals developed different mechanisms to ease the burden of maintaining their social relationships - recognised as one of the most expensive tasks in terms of cognitive resources. For example, primates use grooming to reinforce their alliance with others. Similarly, humans developed language as a convenient and light-weight instrument to maintain their relationships [8]. Language has been refined and evolved in various forms, such as *gossip* - that enables people to maintain more than one social relationship at the same time talking with friends about other friends - and *mentalisation* - the process that allows a person to understand the mental state of other people.

In this context, OSN can be seen as an example of evolution in the language domain. Using them people are able to talk directly to friends or kin, to send broadcast messages to all their contacts, to access information and news, to interact with entities other then humans (e.g., companies, institutions, associations) and to create, access and share multimedia content that can be used to express their feelings, all in one place.

Since the advent of OSN, many different studies have been done to understand the social properties of online social environments, both from a SNA perspective and from an evolutionary point of view. Despite this, we think that there is still a lack of knowledge regarding the extent to which OSN are changing human behaviour in our society and how they will contribute to the evolution of our social relationships in the future.

In this paper we contribute to fill this gap, by analysing the dynamic properties of social relationships in Twitter. This allows us to make well-grounded, though initial, hypotheses on the reasons - from an evolutionary psychology perspective - behind the results that we found. Going more in detail and validating these hypotheses is an extremely interesting subject of further investigation, that will however need custom experiments to verify individual results, and is therefore not covered in this paper.

In the following Section we summarise the most important analyses in literature about online social networks.

2.2 Online social networks

Since the advent of the digital era, humans have introduced new methods to interact within the virtual world of digital communications. The availability of digital communication traces, recorded and stored in centralised servers, paved the way for new opportunities in social network analysis. OSN data have been used to validate hypotheses that were tested on small samples collected through expensive and time consuming questionnaires, or that were simply impossible to test before the advent of OSN due to the lack of data. The typical structural properties of social networks, including the small world effect, have been found in many different OSN (see for example [17, 18]).

Besides the studies regarding the structural properties of OSN, other work has been done to characterise the local properties of social relationships in OSN. For example, in [12], the authors try to define the relationship between the tie strength - the importance of a social relationship in the network - and the different observable variables obtained through OSN communication data. The same authors applied the created model on a different medium, finding consistent results in terms of tie strength among the same social relationships in different OSN [11]. This kind of analysis has been carried out in greater detail in [4], where the authors find a connection between the definition of tie strength given by Granovetter in [15] and the composition of factors - formed of observable variables downloaded from Facebook - that explain the emotional closeness in the online relationships. The evaluation of emotional closeness collected by the authors allowed them to find a first evidence of the presence of the Dunbar's number in Facebook. Another analysis, performed on Twitter, validated the presence of an asymptotic behaviour in the communication patterns ascribable to the idea of the Dunbar's number [13]. In [2] and [3] the authors found ego network structures in Facebook and Twitter similar to those found in offline social networks, with concentric layers with compatible size and scaling factors. In [3], the authors found evidences of a difference in behaviour between separate types of users in Twitter, with those related to "humans" showing a limit ascribable to the Dunbar's number and those that appear not to be "human" who were not affected by such constraints. Recent work on phone call social networks showed that there are some important properties of OSN ascribable to human behaviour. For example, in [24], the authors found that the tie strength is not evenly distributed within ego networks, but it follows a specific shape, called "social signature", characterised by the presence of a few strong ties and many more weak ties - in line with the ego network model described in Section 2.1.2. In [21], the authors give an interesting insight into the dynamics of social relationships in phone call social networks. They identify the presence of a limited capacity each ego can devote to social activity. Moreover, social relationships are dynamically activated and deactivated over time, resulting in a constant ego network size.

Some work has been done to analyse the dynamic aspects of OSN (see for example [14, 30]), but it is mainly focused on the study of the growth of the number of social relationships in the network over time. In this paper we look in detail at the the evolution of the different social structures of the ego networks of the users, that, to the best of our knowledge, has never been done before.

In [30] the authors analysed a large-scale data set obtained from a popular Chinese OSN (RenRen) studying the dynamics of the network. They found that users are most active in building their links shortly after joining the network, eventually decreasing their activity over time. They also found that the presence of communities has a significant influence on user's behaviour. In fact, users belonging to a community are more active in creating new social links, they have longer lifetime and they interact more with other peers in the same community compared to stand-alone users. In [14] the authors built a social network formed of publicly available profiles on Google+, augmenting the network with four additional attributes (i.e., school, major, employer and city) for each node. The results revealed that in some cases the network of attributes shows distributions significantly different from the plain social network. Moreover, attributes have a strong impact on social structure, with interesting differences among different attributes. In [19], a similar analysis on Google+ is presented. The study is focused (as in [14]) on the analysis of statistics about the structure of the network (e.g., clustering coefficient, degree distribution). The results are sometimes in contradiction to the findings in [14] and are focused on the relation between network properties and the geographic distribution of the users, an aspect that is orthogonal to our analysis.

Compared to our analysis, the work presented in [14,19,30] present similar results in terms of growth rate of the number of social relationships over time. Despite this, the analyses are focused on the global growth of the network and they do not consider local aspects of the ego networks of the users. Moreover, they consider only the network of social contacts (i.e., the existence of social links), without weighting the links by the interactions occurred between people. This is clearly an important aspect that must be considered to correctly analyse the social behaviour of people in online environments, even though, to the best of our knowledge there are no large-scale data sets about OSN other than Twitter with the same detailed information needed for this kind of analysis. Specifically, it is difficult to obtain the whole communication history of the users with timestamps about the single communication events needed to track the dynamics of ego networks.

Recently, a study on Twitter [23] revealed that the structure of the ego networks of the users is related, as in offline environments, to parts of one's social world (i.e., to topics, geography and emotions). The authors collected the last 200 tweets of a large sample of Twitter users (about $250,000$) and, after a detailed categorisation of the messages sent by the users into different topics, they found that users with less-constrained ego networks structures (i.e., with access to structural holes) tend to cover diverse topics. Moreover, their results highlight that the majority of the users have geographically-constrained networks and that the users are clustered according to happiness (calculated with a sentiment analysis on the communication traces). Compared to [23], we have been able to download much more data for each user (up to $3,200$ tweets), that allowed us to make a detailed analysis about the evolution of the social ego networks.

Although the work done so far on the analysis of OSN has identified some important aspects of human behaviour (see [1] for a survey on the aspects of human behaviour identified in Twitter), the dynamics of the processes governing human social behaviour in online environments are still unknown. In this paper we make a contribution to the field with a fine grained analysis of the evolution of ego networks over time in Twitter. This analysis reveals many important aspects of human behaviour in OSN.

Before introducing the analysis, we describe the data set we have collected and studied.

3. DATA SET DESCRIPTION

In this Section we describe the data set we used for the analysis. We describe the crawler we used to obtain the data from Twitter, the classifier we used to select profiles related to "humans" and the descriptive statistics of the data. Before continuing with the description of the data set, we briefly introduce Twitter and the different communication mechanisms it offers to its users.

3.1 Twitter and tweets

Twitter is an online social networking service founded in 2006 that has reached a very high popularity, with more than 500 million registered users as of 2012[1]. The main feature of Twitter is to enable its users to send short text messages called *tweets*. Tweets can be up to 140 character-long messages containing tags to reference other users, keywords to identify the topic of the message and links to web pages or multimedia content.

Users can *follow* other users to automatically receive all their tweets and visualise them in their home page. The users that a person follows are her *friends*, whereas people that follow that person are her *followers*. The act of referencing a user in a tweet is called *mention*. Mentions are direct messages sent to one or more people through the mention mechanism and are a special form of direct communication between users. Twitter enables users to directly *reply* to any tweet automatically adding a mention to the response. Replies often involve bi-directionality in the communication, since they are mostly used to reply to previously received mentions. Twitter allows the exchange of private messages as an additional mechanism for direct communications. Despite this, the content of these messages is private and cannot be accessed without the user's permission. Moreover, private messages represents only a small subset of all the messages exchanged on Twitter and therefore using only them to identify direct communications between users may result in an incomplete picture. For these reason, we do not consider them in our analysis and in the rest of the paper we refer to "direct communication" as the public direct tweets only (i.e., mention and replies). Besides direct communication, all the tweets are automatically broadcast to all the users' followers. Tweets can be *retweeted* or, in other words, forwarded by users to all their followers. Retweeting is a really efficient communication means to rapidly spread information in the network. The special tags used to assign one or more topics of a tweet are called hashtags, since they are characterised by the presence of the "#" character before the name of the topic, as part of the text of the tweet. Hashtags are used by Twitter to classify tweets and to cluster them into categories, browsable by the users.

3.2 Data download process

We downloaded and analysed a large sample of Twitter communication data which we used to build ego networks representing personal online social relationships of Twitter users. We collected data about Twitter user profiles and the complete history of their tweets. To obtain the data we crawled Twitter from November 2012 to March 2013, downloading a total of $2,428,647$ complete user profiles. The crawler agent we used is described in more detail in [3]. The crawler uses Twitter REST API to collect data regarding

[1]According to Twitter CEO Dick Costolo in October 2012

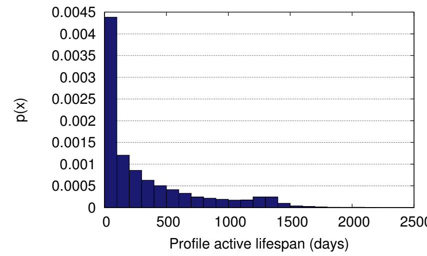

Figure 2: Distribution of active lifespan of Twitter ego networks

user profiles and all their tweets, for a maximum of $3,200$ tweets per user - due to the restrictions imposed by the API. The crawler follows links between users to build a network of connected profiles[2]. Specifically, it uses the following/followers lists and the content of direct messages (i.e., replies and mentions) to identify new profiles to download. The data we obtained is an extension of the data set used in [3]. The present data set contains many more profiles, since we ran the crawler for four additional months. This allowed us to carry out a detailed analysis of the dynamics of Twitter ego networks.

3.3 Detecting humans in the crowd

Since our goal is to analyse human behaviour in Twitter, we isolated from the data set the user profiles presenting recognisable human characteristics, discarding all the other kinds of profiles that do not appear to be "humans". For example, we want to discard profiles run by companies, institutions, bots and all the other profiles that can be intuitively classified as "non-humans". Results in [3] showed that this intuitive distinction is accurate enough to separate users characterised by well-known social properties ascribable to humans (i.e., limits imposed by cognitive constraints) from users without these properties, evidently not humans. To automatically perform the separation between "humans" and "non-humans" we used a classifier, specifically a Support Vector Machine (SVM), already used in [3]. The method we used is similar to that proposed in [6], where the authors present a supervised learning approach to classify Twitter account into four different categories (i.e., organisations, journalists/bloggers, ordinary individuals and other). Our SVM uses 96 features extracted from Twitter data. These features are related both to users' profiles and their tweets (e.g., number of friends and followers, whether the profile has been verified by Twitter or not, number of tweets, retweets, replies and mentions sent by the user). We trained the SVM with 500 manually classified user accounts. The manual classification was carried out taking a random sample of profiles from Twitter and giving a binary classification based on the visual inspection of each Twitter profile page. After training the SVM we tested its accuracy on a test set of 100 accounts, which we manually classified, but that were not used during the training phase. The accuracy of the SVM is 0.813 with a 95% c.i. equal to (0.789, 0.837). Although this result could be improved using a larger training set, it is comparable with the results in [6] and is sufficient to draw significant results about human behaviour

[2]The crawl started from a very popular user, so that we could immediately have a large sample of other users at the first hop of the process.

in Twitter, while required to analyse only a very small percentage of crawled accounts.

3.4 Data set properties

After applying the SVM on the data set we obtained 1,653,155 "human" profiles, about 68% of the total number of users in the data set. To the best of our knowledge, this is the first time the estimation of the percentage of "human" profiles is calculated on Twitter. The large number of "non-human" profiles in the data set gives us a first interesting picture of Twitter. In fact, it indicates that Twitter is an online environment where different types of users coexist and interact. This feature makes OSN different from more traditional communication means, which often create a separation between different social environments. People using Twitter receive multiple social benefits at the same time, being able to manage more social domains in the same place.

In the first column of Table 1 we summarise the properties of the profiles in the data set, considering human users only. We report the mean values of the indicated statistics, averaged for all the profiles. The active lifespan of the profiles, the distribution of which is depicted in Figure 2, indicates the temporal length of the period in which each profile actively sent tweets. The active lifespan of a user starts with her first tweet and ends with her last tweet. Since our analysis is focused on the social activity of the users on Twitter, we think that measuring the duration of a Twitter account using its active lifespan is appropriate for this work. Of course, some of the users in Twitter could be lurkers, not actively investing their resources in the maintenance of their social relationships. Since the aim of this study is to obtain a detailed characterisation of the evolution of the social structures actively maintained in Twitter through direct communications, we are not interested in lurking-only users, for they do not actively interact with other people. Moreover, as pointed out in [22], there is not a sharp separation between lurkers and active users, since all the users alternate between lurking behaviour and active behaviour when using social-oriented virtual environments. For this reason, we think that considering direct communications is sufficient to capture the social behaviour of all the different types of users in Twitter.

To make the definition of active lifespan more clear, let us consider the following example. If a profile had been created four years before the download, but the user associated with the profile sent tweets only during the first year and then stopped using Twitter (at least for sending tweets), the resulting active lifespan of the profile is one year. The shape of the distribution in Figure 2 indicates that either most of the profiles have been created just before we downloaded the data or their activity on Twitter is very low. However, the long tail indicates that we were able to obtain profiles with a

Table 1: Data Set Statistics

variable	mean - all	mean - active
duration (days)	321.846 [0.628]	448.201 [0.762]
replies	208.923 [0.609]	290.885 [0.801]
mentions	103.882 [0.459]	144.634 [0.625]
retweets	151.492 [0.496]	210.924 [0.661]
plain text twts	280.037 [0.773]	389.810 [1.011]
twts w urls	4.813 [0.032]	6.698 [0.045]
twts w hashtags	56.411 [0.203]	78.529 [0.273]

tweet history of up to almost 7 years (i.e., the complete tweet history of some of the oldest profiles in Twitter), despite the limit of 3,200 tweets imposed by the Twitter API. During the manual classification of the training set (described in Section 3.3), we noted that "humans" rarely generate more than 3,200 tweets, and most of the profiles exceeding the limit were "non-humans"[3]. Indeed, only 0.02% of the "human" profiles in the data set exceed this limit. Nevertheless, the small peak in the distribution between 1,200 and 1,400 days could be ascribed to the presence of this limit, that prevented us from obtaining the complete active lifespan of some of the downloaded profiles. Despite this, the number of profiles affected by this problem is very low and their last 3,200 tweets are in any case a significant sample to describe their social behaviour. For this reason, the data set we collected is well suited for our analysis.

In Table 1, we notice that the mean active lifespan of the "human" profiles in the data set (i.e., "duration" in the Table) is equal to 321.846 days. This indicates that, on average, we captured almost one year of communications for each user and this is sufficient to conduct our analysis. Replies and mentions are about 39% of the total number of tweets made by "humans" in Twitter. The exchange of these messages can be interpreted as a mechanism to actively maintain social relationships online and they should be strongly affected by our cognitive limits, since they require the users to spend cognitive resources to directly communicate with the involved people. Besides, non-direct messages take the largest part of the communication in Twitter. This kind of communication is controlled by a more *public* behaviour compared to direct messages and it should require less cognitive resources, since we expect non-direct tweets to contain a low value of emotional intensity.

The high number of replies could indicate a high number of communication threads between people. In fact, replies are usually used to reply to a previous mention and communication threads are composed by an initial mention and a series of replies to that mention. The presence of communication threads is supported by the fact that the number of replies is, on average, broadly twice the number of mentions. This is another strong indication of the maintenance of social relationships online. Retweets are largely used by Twitter users and represent the willingness of people to spread messages they are interested in within the network. Seen from an evolutionary perspective, the diffuse usage of retweets could represent a strategy used by humans to receive a global benefit from having access to more information in the network, at the cost of being active in the diffusion process.

Remarkably, non-direct tweets containing urls are less used than the other type of messages (only 4.813 tweets with urls sent on average by the users in our data set during their active lifespan). In addition, the low number of tweets with hashtags (i.e., 56.411 on average) could be ascribed to the fact that Twitter officially introduced hashtags only between 2009 and 2010.

After selecting humans from the data set, we discarded all the profiles that have not sent any tweets (i.e., with null active lifespan), reducing the number of profiles to 1,187,105.

[3]Even though the SVM is not only based on this feature and is able to correctly classify cases of "non-humans" tweeting less than 3,200 times in a period of time compatible with "human" profiles

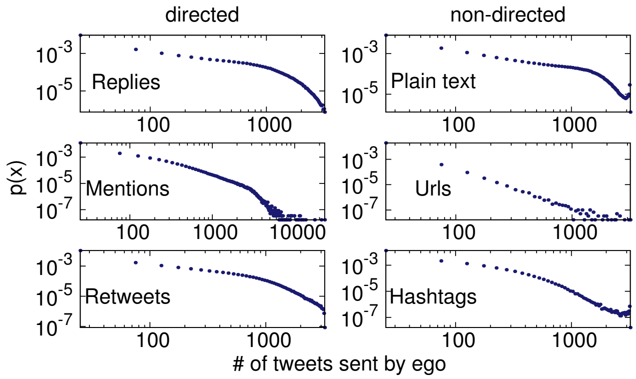

Figure 3: Distribution of the number of tweets divided by type

The statistics of these set of profiles are reported in the second column of Table 1. We can notice that all the statistics of the profiles increase when users with null active lifespan are not considered, since the removed profiles do not contribute actively to the generation of content in the network.

Figure 3 depicts the distribution of the communication variables in the data set for the profiles with positive lifespan. We separated direct and non-direct communication, with the former identifying the explicit intention of the user to mention other users in the messages. In the figure we labelled retweets as direct communication, but their nature needs further investigation. In fact, retweets are more similar to non-direct tweets, with the exception that they contain the id of the user that initially generated the message and the ids of users that retweeted it.

Mentions show a very long tail (the scale of the x axis is different than for the other graphs), with some accounts generating up to $23,104$ mentions. This high number of mentions - apparently exceeding the limit of $3,200$ tweets - is due to the fact that a single tweet can contain more than one mention at the same time, or, in other words, many people can be mentioned in the same tweet.

Before continuing with the analysis we further filtered the data set, eliminating all the profiles created less than one year before the time of their download. This reduces possible artefacts due to including recently created accounts (with respect to the end of the crawl) as well as accounts that have been active only for a short amount of time. The data set, after this selection, contains $644,014$ accounts.

3.5 From tweets to ego networks

From the set of active human users, we built a social ego network for each profile. To do so we firstly defined a measure of the strength of social links between people in Twitter. We say that a social relationship exists between two users, A and B, if A sent at least a reply or a mention to B. This definition involves a cost in terms of cognitive effort spent for the maintenance of the relationship. As an estimate of the tie strength we use the number of messages sent by A to B. In this way, tie strength grows linearly with the number of messages exchanged between two users. We think that representing tie strength in this way is, at the moment, the best possible solution, since models to study the relation between tie strength and frequency of contact are still under investigation in OSN, and the first results indicate that using linear approximations leads to sufficiently accurate results [4].

Using the standard terms in ego network analysis, we call *ego* a user associated with a profile and *alters* all the people with whom ego has a social relationship. This definition gives a "static" view of the ego networks in the data set, aggregating all the communication of the egos, as typically done in other studies [2–4,13,16]. This allowed us to make a qualitative comparison between the ego network size in our data set and that found in other studies in the literature, before moving to analysing the dynamic properties of social relationships. The total number of social relationships in the data set is $57,548,091$ with an average of 89.36 relationships per profile. This result is in accordance with the findings of other studies on OSN [2–4,13], but is considerably lower than the Dunbar's number found in offline ego networks [31]. This could be due to the fact that Twitter is only one of the many possible tools used to maintain social relationships and the time dedicated to socialising in Twitter is still limited [7].

To better understand how these ego networks evolve over time we analysed the time series of the tweets sent by ego and we studied the composition of snapshots of the ego networks considering the communication occurred in time windows of one year each. This allowed us to reveal important insights regarding human social behaviour in OSN.

4. METHODS

To perform the analysis, we studied the time series of the direct tweets (replies and mentions) and of the non-direct tweets sent by each ego. For some performance indices (i.e., new users contacted per day and total number of new users contacted) we counted the number of new alters contacted by ego each day until the network is active. Instead, for analysing the dynamics of the ego network structure, we sliced the tweets time series taking snapshots of the duration of one year each, then assessing the size and the composition of ego networks in each snapshot. We slid the one-year temporal window taking steps of one day each, looking at how ego networks change over time.

By taking temporal windows of one year we were able to capture all the active contacts maintained by each ego and their evolution over time, according to the definition of active network introduced in Section 2.1.2 that identifies as "active" friends all the alters contacted by ego at least yearly. In this way we were also able to identify relationships that the users abandoned over time. Note that we do not use the notion of "unfollowing" (i.e., the explicit request of a user to remove a person from her friends) to identify abandoned relationships, since unfollowing is an extreme action that does not capture the decline of a social link, but rather identifies sudden breach in the relationships, due to particularly negative and rare conditions.

We defined as sympathy group the set of alters contacted at least once a month (i.e., contacted at least ~ 12.17 times in one year), and as support clique the alters contacted once a week (i.e., contacted ~ 52.14 times in one year). Doing so, we were able to analyse how the different layers of the ego networks change over time. We refer the reader to Section 2.1.2 for the definitions of the different ego network layers.

To be able to analyse the average behaviour of all the ego networks we shifted the first communication of each ego network (the time when ego started to actively communicate), so that they start at the same point in time, specifically at the origin of the coordinate system of each figure reported in the following Sections.

To deeply analyse the behaviour of different users in Twitter, we divided the users in three categories on the basis of their active lifespan and we studied the differences in terms of social behaviour between these classes. To do so, we took the maximum lifespan in the data set and we divided it into three equal parts, obtaining three groups of 802 days of duration each. We decided to create exactly three categories since this choice represents a good trade-off between the accuracy of the results and their statistical significance. In fact, adding more categories would have decreased the number of users in each group, leading to low significance. After the categorisation we defined the following classes of users: (i) occasional users (lifespan $<= 802d$) ; (ii) regular users ($802d <$ lifespan $<= 1604d$) and (iii) aficionados (lifespan $> 1604d$). We expect these different categories of users to show different behaviours and different ego network properties. Our data set is composed of 63.23% of occasional users, 35.22% of regular users and 1.55% of aficionados[4].

Note that in the figures presented in Section 5, regarding the composition of ego networks in each one year snapshots (right-hand side plots in Figures 4 to 6 and 9), the value of the x axis represents the starting point of each snapshot. Thus, the maximum value of the axis is equal to the maximum lifespan of the ego networks in the considered class, minus the duration of the snapshot (one year). In the figures we report the average values as the curve in bold and the corresponding 95% confidence interval as a lighter coloured area around the curve (barely visible, most of the time).

As another contribution of this paper, we analysed the evolution of the recency of contact (i.e., time since last contact) between users, to understand how single social relationships evolve. To do so we measured the elapsed time between consecutive messages within each relationship. We averaged the results within the ego networks and then averaged for all the ego networks. While this clearly mixes the properties of different type of social relationships for a particular ego network, it provides a unique index that allows us to compare the ego networks of different classes of users, as explained in detail in Section 5.

After the analysis of the evolution of ego networks and personal social relationships over time, we measured the stability of ego networks, assessing the proportion of alters that users maintain in their networks over time. We estimated this proportion by comparing consecutive - but separated - one year snapshots and calculating their average Jaccard coefficient, then averaging the results for all the ego networks. The Jaccard coefficient is a measure of the percentage of overlap between sets defined as:

$$J(W_1, W_2) = \frac{|W_1 \cap W_2|}{|W_1 \cup W_2|} \quad (1)$$

where W_1 and W_2 are two sets, in our case the one-year windows of the ego networks. The Jaccard coefficient can be a value between 0 and 1, with 0 indicating null overlap and 1 a complete overlap between the sets. We calculated the Jaccard coefficient for the different layers in the ego networks. This allowed us to determine the "turnover" that takes place in the ego networks. This study is fundamental for understanding whether people maintain a stable network of contacts in Twitter or they prefer to vary their social relationships over time, and allowed us to define two distinct classes of users: (i) users with structured ego networks, showing ego networks with composition and turnover similar to those found in other more traditional social networks and (ii) people without structured ego networks, showing higher turnover.

[4]Note anyway that we still have around 10,000 aficionados in our data set, which makes the analysis of also this class significant

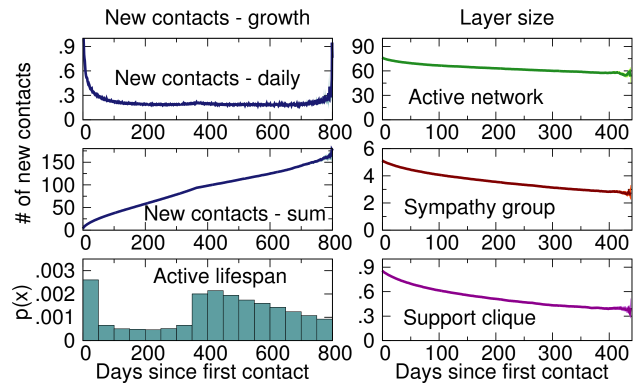

Figure 4: Ego networks properties for occasional users

5. RESULTS

In this Section we report the results of our analysis and we interpret them from the point of view of human social behaviour. The main axes of our analysis, as identified in Section 4, are the presence of different categories of users and, on the other hand, the presence/absence of a structured ego network.

5.1 Twitter abandonment

As a first contribution of our analysis, we studied the behaviour of users that abandoned Twitter. We say that a user has abandoned Twitter if her active lifespan is followed by a period of at least six months of inactivity. In the data set, the average active lifespan of users that abandoned Twitter is 73.21 days, indicating that most of them are occasional users. In fact, over a total of 159,069 accounts that abandoned Twitter (i.e., 24.7% of our data set), 88.27% are occasional users, whilst only 11.6% are regular users and 0.13% are aficionados. From the distribution of the active lifespan of occasional users (depicted in the bottom left part of Figure 4) we can notice that there is a small number of accounts with duration between 50 and 365 days. Yet, there is a non negligible number of occasional users with a very short lifespan (i.e., $< 50d$). These accounts represent people that joined Twitter more than one year before the download, but that abandoned it after a short period of activity. This class of users can be seen as a sub-class of occasional users, who subscribed to Twitter only to "give it a try", but abandoned it very soon.

5.2 Ego networks evolution over time

5.2.1 *Number of different alters contacted*

The first result worth mentioning is that the number of new people that egos contact grows at a constant rate. This is true for all the categories of users and can be seen in the top left graphs in Figures 4 to 6. The graph labelled "New

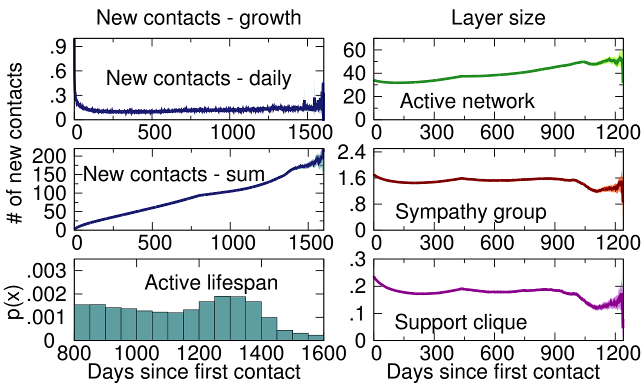

Figure 5: Ego network properties for regular users

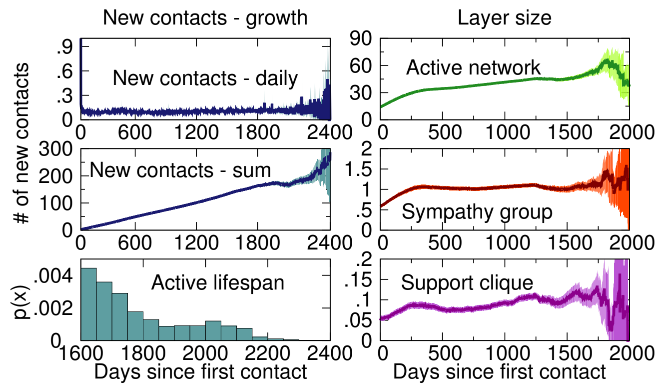

Figure 6: Ego network properties for aficionados

contacts - daily" depicts the number of new users contacted by egos during each day of their activity (averaged over all users still active at that day), whilst "New contacts - sum" represents the cumulative number of new users contacted by ego over time (again, averaged across all active users). From these graphs it is clear that, after a first phase in which ego contacts new people at a higher rate, this number quickly converges to a constant. The value of this constant is higher for occasional users than for the other classes. The mean over time is 0.222, 0.125 and 0.112 for the three classes, respectively. This indicates that occasional users have more dynamic ego networks, with a higher number of new social links added over time compared to the other categories. We can notice that the total number of different people contacted by egos over time is, on average about 200 and it is constantly growing, with little variation between the different classes, even though the duration of the ego networks changes considerably between classes. These results are in accordance with the findings in [30], where the authors found that users in RenRen (a popular Chinese OSN) are more active in creating new social links shortly after joining the network. The users eventually approach a constant number of edges created per time unit once most offline friends have been found and linked.

The presence of a constant growth rate is an important aspect of human social behaviour, indicating high dynamism in the ego networks of the users, that are constantly contacting new people rather than maintaining a limited number of stable relationships. This behaviour is confirmed by the analysis of the set of people actively contacted within the ego networks, reported below.

To understand how the constant addition of new contacts in the ego networks impacts on the communication level with the set of existing alters, we studied the evolution of the size of the set of alters actively maintained over time, as reported in Section 5.2.2. Moreover, in Section 5.5 we report the analysis of the percentage of turnover (i.e., the degree of variation in the set of alters actively contacted) for the different layers in the ego networks.

5.2.2 *Number of alters actively contacted*

Even though the number of new alters contacted by egos increases over time, the number of alters that are actively maintained in the ego networks does not increase at the same rate. This fact reveals the presence of a turnover strategy within the ego networks, since the new contacts replace other relationships that are not maintained by ego. The size of the ego network layers are depicted in the right column of Figures 4 to 6, for the different categories of users. As far as occasional users are concerned, the size of all the layers significantly decreases over time. Specifically, the active network has a total decrease of 30.73%, the sympathy group of 45.91% and the support clique of 53.22%. Regular users show a different behaviour, with a considerable increase in the active network size (31.16% in almost 4 years), but with a decrease in the other layers (32.17% for the sympathy group and 30.42% for the support clique). It is worth noting that occasional users, compared to regular users, show a higher value of new contacts added in their ego networks daily and larger sizes in all the layers at the beginning of their lifespan, eventually approaching sizes compatible with the regular users. Aficionados show a considerable growth in size in all the ego network layers, even though the rate at which they contact new people is lower than for the other categories. These results highlight the different behaviour of the users in Twitter and indicate that occasional users have an initial boost of activity followed by a decrease or a sudden abandonment of the platform. Regular users and aficionados have a slower start, but they eventually increase the size of their active network over time. Aficionados even increase the size of their inner layers, indicating an investment in strong social relationships, maybe due to the longevity of such relationships, constantly reinforced through Twitter.

On average, the active network size lies between 30 and 80 for all the categories. This result suggests the effect of cognitive constraints of human brain in online environments, which limit the number of people that can be actively maintained over time, in line with the concept of the Dunbar's number. The small active network size, compared to offline social networks size found offline (equal to 132.5 [31]) can be related to the fact that Twitter is only a part of the complete social network of the users and the time spent on Twitter is still low compared to the time spent socialising in person, even though this discrepancy is constantly decreasing [7].

The lower growth rate shown by the sympathy group and the support clique compared to the active network (even negative for occasional users and regular users) suggests the presence of a strategy whereby people prefer dynamic ego networks formed of light-weight social relationships that give access to a larger amount of network resources [15], rather than more stable ego networks with stronger and well-consolidated relationships. Note however that for aficionados (i.e., users that spend a lot of time maintaining their social relationships in Twitter) this preference towards

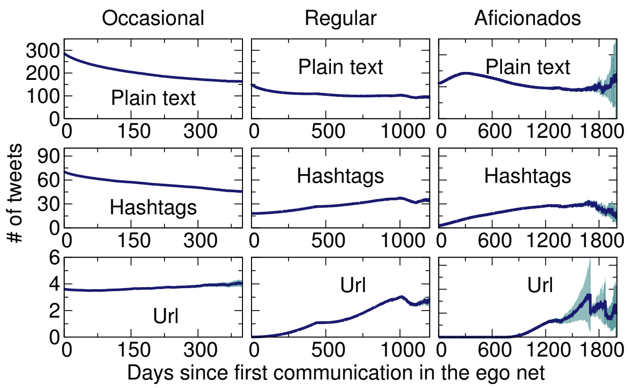

Figure 7: Non-direct communication divided by category

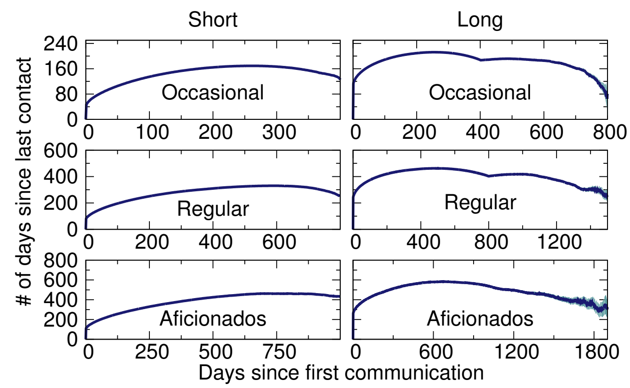

Figure 8: Days since last contact evolution over time

light-weight social relationships is way less marked, making their behaviour much more similar to the one highlighted in previous studies of social networks [31].

Finally, the rate at which egos contact new users is negatively correlated with ego networks growth rate, indicating that users spending a lot of their time adding new people to their networks do not have enough resources to maintain all these relationships over time and their layers inevitably decrease in size. This is in accordance with the idea that our social capacity is limited by cognitive constraints and going beyond our limits could even brake up our social network [25].

5.3 Non-direct communications

We studied how the number of non-direct tweets (i.e., plain text tweets, tweets with hashtags and tweets with urls) change over time for the different categories. The results are depicted in Figure 7. Occasional users significantly decrease the amount of non-direct tweets they send over time - apart from tweets with urls, although these are very limited. This category of users shows an initial boost of activity followed by a gradual decrease, as already found for direct communications. Regular users show a much more stable trend for what concerns the number of plain text tweets, with a value asymptotically converging towards ~ 100 tweets sent in each one-year window. Yet, the number of non-direct tweets is noticeably lower than for the previous category, even though it is increasing over time. This indicates that regular users are less affected by an initial boost, and they rather have a slow start. Aficionados show a similar pattern, apart from plain text tweets, which show a peak in the first two years of their active lifespan. This peak could be due to an initial enthusiasm in the platform at a global level, since this category contains some of the oldest profiles in Twitter. After this initial phase, the number of plain text tweets converges asymptotically to a value similar to the other classes.

These results tell us that whilst some users abandon Twitter after a short period of time, the activity of the egos that continue to use the platform remain stable, rejecting the hypothesis of a convergence towards the OSN decline [26, 29]. This is in contrast with the results of [28], where the authors found that, in Facebook, users are more active when they join the network, decreasing their use rate over time. Our analysis reveals that this behaviour is true only for occasional users and that there is a non negligible amount of long-term users contributing to the survival of the OSN.

5.4 Evolution of personal social relationships

To better understand how personal social relationships evolve in Twitter, we analysed how the average time since last contact changes over time for each single social link in the different categories. We divided the social relationships in each category into "short" relationships, with duration shorter than half of the maximum duration of the category, "long" relationships, with duration longer than the same threshold. Figure 8 depicts the number of days since last contact between people involved in each social relationship (on the y axis) as a function of the time since the beginning of the relationship (x axis). From the figure we can notice that all the distributions show a "bow" shaped curve. This particular shape tells us that, on the one hand, social relationships have an initial phase in which they have a shorter time since last contact (i.e., higher frequency of contact) followed by a gradual increment. On the other hand, since some social relationships disappear as time passes, the remaining social relationships have shorter time since last contact, resulting in the gradual decay in the right most part of the graphs.

It is worth noting that there is a significant variation in the values of time since last contact in the different categories of users, with occasional users having lower values compared to the other classes. Once again, this supports the idea for which occasional users have an initial boost of activity, followed by abandonment or gradual decay.

5.5 Ego network turnover

Finally, we assessed the stability over time of each layer for the different categories. To do so, we calculated the average Jaccard coefficient between separated one-year windows in each ego network. To perform this analysis we further reduced the number of ego networks in the data set, since we needed at least two years of active lifespan to calculate the Jaccard coefficient between two different non overlapping one-year windows. Thus, we selected $190,249$ ego networks with active lifespan greater than two years. The average Jaccard coefficients for the different layers are reported in Table 2 under the label "all ego networks". The low values of Jaccard coefficient for all the layers indicate a percentage of turnover higher than 75%, with a maximum of 98.8% for the support clique of aficionados. This reveals that the average turnover in each layer is really high. Interestingly, the turnover in the inner layers is higher than the turnover in the active network. This result is in contrast with the findings on phone call records analysed in [24], where the

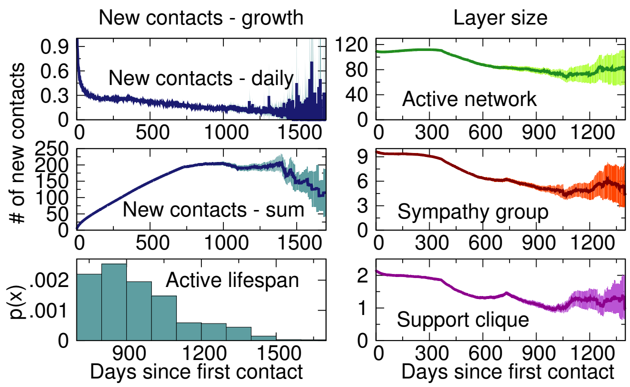

Figure 9: Ego network properties of structured ego networks

authors found that for the top 20 ranking alters in ego networks - formed of social links weighted with the number of calls between people in a fixed time period - the turnover is lower than for the rest of the ego network. It is also worth noting that occasional users show higher stability compared to the other classes. This result could be explained by the fact that the longer the lifespan, the higher is the probability that the social relationships in the ego network change due to turnover.

The low values of Jaccard coefficient in the inner layers (i.e., 0.057 for occasional users, 0.024 for regular users and 0.012 for aficionados) could be influenced by the presence of small support cliques and sympathy groups, that for many egos do not even exist. For this reason we decided to calculate the Jaccard coefficients considering only users that always maintain a structured ego network, or, in other words, that show a non empty support clique in all the sampled one-year windows. The results are reported in Table 2 under the label "structured ego network". In this case the values on the Jaccard coefficient for the different layers are higher than in the previous case and are compatible with the findings in [24]. The values of the percentage of turnover of the active networks are similar for all the different categories and are about 81% (Jaccard coefficient ~ 0.19). For what concerns the other layers, the sympathy group show a percentage of turnover between 71.3% and 63.8%, whereas the support clique 65.4% and 51.2%. These results denote a behaviour similar to other social networks, where the inner layers contain stronger relationships that should be intuitively less affected by the turnover in the network. Nevertheless, as already found in [24], also the inner layers are strongly affected by turnover. The number of ego networks that show a turnover pattern similar to those found in other social environments is 10,307, only 5.42% of the analysed egos. This is another strong indication that human behaviour in Twitter significantly differs from other social networks involving more traditional and dyadic communications. Remarkably, in structured ego networks the categories of users with longer lifespan have higher values of Jaccard coefficient, especially for the inner layers. This tells us that users that maintain structured ego network tend to reinforce their close relationships over time, instead of devoting their time to supporting weak relationships. Note that this is in accordance with the analysis of the evolution of the sizes of the layers over time for aficionados, discussed in Section 5.2.

Table 2: Average Jaccard coefficient of different network layers

layer	Occasional	Regular	Aficionados
All ego networks			
active net	0.124	0.098	0.103
sympathy gr.	0.122	0.075	0.072
support cl.	0.057	0.024	0.012
Structured ego networks			
active net	0.191	0.190	0.193
sympathy gr.	0.287	0.309	0.362
support cl.	0.346	0.395	0.488

We have further analysed the properties of these 10,307 ego networks applying the same technique used in Section 5.2.2. The results are shown in Figure 9. The active lifespan of these ego networks ranges between 730 and 1,749 days. These are the minimum and maximum active lifetimes of ego networks in our dataset that always presented a non-empty support clique. This definition allowed us to isolate users with behaviour similar to that showed in "offline" environments, where the support clique is maintained over time by the majority of people as the most important part of their networks.

Interestingly, the layers of the structured ego networks are larger than the average, resembling the layers found in [3], where the authors identify in Twitter a "super support clique", as a set containing one or two alters with very strong relationships with ego, perhaps a partner and/or a best friend. Also the sympathy group and the active network sizes are compatible with this previous study. Remarkably, all the layers decrease in size as time passes and so does the number of new alters contacted by ego. This could be explained by the presence of the initial boost of social activity of occasional users. Nevertheless, egos with longer lifespans prefer to consolidate their social relationships than adding new contacts, as indicated by the decrease in the top left graphs in Figure 9. This is in accordance with the results presented in the previous Sections.

6. CONCLUSIONS AND FUTURE WORK

In this paper we presented a detailed analysis of the dynamic processes of ego networks and personal social relationships in Twitter. The results indicate that human behaviour in Twitter significantly differs from other social networks studied in literature in different research fields. On average, compared to more traditional social networks, Twitter presents smaller ego networks with a high percentage of weak ties and a really high turnover. This fact led us to the conclusion that the general behaviour of Twitter users is to maintain a light-weight ego network formed of weak social relationships suitable to maximise the amount of resources accessible trough the network and limiting the number of strong relationships. This type of user shows an initial phase of very high activity that is inevitably followed by a gradual decay or abandonment. On the other hand, a small but noticeable set of users prefer a "slow" start with a gradual increase of activity and more stable networks. This type of user shows ego networks much more similar to those found in previous analyses of social networks, with more stable inner layers and larger active networks (with respect to the first type of users). Moreover, our results also indicate that users that do not immediately abandon Twitter tend to use it at a regular rate in terms of direct and non-direct communication. This suggests that the hypothesised decline in the use of OSN might not be present, at least in Twitter.

Seen from an evolutionary perspective, the presence of a vast majority of users of the first type, and the resulting difference between the properties of their Twitter networks and conventional models of ego networks represents an interesting fact, since their behaviour seems to be adapting to the dynamism of our society, reflected in the need of new ways of acquiring information in a very dynamic way through OSN like Twitter.

Other interesting directions we are exploring at the moment include the study of the relation between ego network dynamics and the structural properties of the social network. This study could represent a further step for bridging the gap between online social network analysis and more traditional approaches derived from social sciences.

7. ACKNOWLEDGEMENTS

This work was partially funded by the European Commission under the SCAMPI (FP7-FIRE 258414), RECOGNITION (FP7 FET-AWARENESS 257756), EINS (FP7-FIRE 288021) and EIT ICT Labs MONC (Business Plan 2013) projects.

8. REFERENCES

[1] O. Aarts, P.-P. van Maanen, T. Ouboter, and J. M. Schraagen. Online Social Behavior in Twitter: A Literature Review. *ICDMW*, pages 739–746, 2012.

[2] V. Arnaboldi, M. Conti, A. Passarella, and F. Pezzoni. Analysis of Ego Network Structure in Online Social Networks. In *SocialCom*, pages 31–40, 2012.

[3] V. Arnaboldi, M. Conti, A. Passarella, and F. Pezzoni. Ego Networks in Twitter: an Experimental Analysis. In *Netscicom*, 2013.

[4] V. Arnaboldi, A. Guazzini, and A. Passarella. Egocentric Online Social Networks: Analysis of Key Features and Prediction of Tie Strength in Facebook. *Computer Communications*, 36(10-11):1130–1144, 2013.

[5] R. S. Burt. *Structural Holes versus Network Closure as Social Capital*. 2001.

[6] M. De Choudhury, N. Diakopoulos, and M. Naaman. Unfolding the event landscape on twitter: classification and exploration of user categories. In *CSCW*, pages 241–244, 2012.

[7] J. Delaney, N. Salminen, and E. Lee. Time americans spend per month on social media sites - sociallyawareblog.com, 2012.

[8] R. Dunbar. Theory of mind and the evolution of language. In *Approaches to the Evolution of Language*, chapter 6, pages 92–110. 1998.

[9] R. I. M. Dunbar. The social brain hypothesis. *Evolutionary Anthropology*, 6(5):178–190, 1998.

[10] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning about a highly connected world*. 2010.

[11] E. Gilbert. Predicting tie strength in a new medium. In *CSCW*, pages 1047–1056, 2012.

[12] E. Gilbert and K. Karahalios. Predicting tie strength with social media. In *CHI*, pages 211–220, 2009.

[13] B. Gonçalves, N. Perra, and A. Vespignani. Modeling users' activity on twitter networks: validation of Dunbar's number. *PloS one*, 6(8):e22656, 2011.

[14] N. Z. Gong, W. Xu, L. Huang, P. Mittal, E. Stefanov, V. Sekar, and D. Song. Evolution of Social-Attribute Networks: Measurements, Modeling, and Implications using Google+. In *IMC*, pages 131–144, 2012.

[15] M. S. Granovetter. The Strength of Weak Ties. *The American Journal of Sociology*, 78(6):1360–1380, 1973.

[16] R. A. Hill and R. I. M. Dunbar. Social network size in humans. *Human Nature*, 14(1):53–72, 2003.

[17] R. Kumar, J. Novak, and A. Tomkins. Structure and evolution of online social networks. *KDD*, pages 611–617, 2006.

[18] J. Leskovec and E. Horvitz. Planetary-Scale Views on an Instant-Messaging Network. Technical report, 2007.

[19] G. Magno and G. Comarela. New kid on the block: Exploring the google+ social graph. In *Proceedings of the ...*, pages 159–170, 2012.

[20] P. V. Marsden and K. E. Campbell. Measuring Tie Strength. *Social Forces*, 63(2):482–501, 1984.

[21] G. Miritello, R. Lara, M. Cebrian, and E. Moro. Limited communication capacity unveils strategies for human interaction. *Scientific Reports*, 3:1–7, June 2013.

[22] M. Muller, D. R. Millen, N. S. Shami, and J. Feinberg. We are all Lurkers: Toward a Lurker Research Agenda. In *CSCW*, pages 1–10, 2010.

[23] D. Quercia, L. Capra, and J. Crowcroft. The social world of twitter: Topics, geography, and emotions. *... AAAI Conference on Weblogs and Social ...*, (Hansen 1999), 2012.

[24] J. Saramaki, E. Leicht, E. Lopez, S. Robetrs, F. Reed-Tsochas, and R. Dunbar. The persistence of social signatures in human communication. *arXiv preprint arXiv*, pages 1–16, 2012.

[25] A. Sutcliffe, R. Dunbar, J. Binder, and H. Arrow. Relationships and the social brain: Integrating psychological and evolutionary perspectives. *British journal of psychology*, 103(2):149–68, 2012.

[26] M. Sweney. Facebook sees first dip in UK users - guardian.co.uk, 2008.

[27] J. Travers and S. Milgram. An Experimental Study of the Small World Problem. *Sociometry*, 32(4):425, 1969.

[28] C. Wilson, A. Sala, K. P. N. Puttaswamy, and B. Y. Zhao. Beyond Social Graphs: User interactions in online social networks and their implications. *ACM Transactions on the Web*, 6(4):1–31, Nov. 2012.

[29] B. Worthen. Bill Gates quits Facebook - Wall St. Journal Online, 2008.

[30] X. Zhao, A. Sala, C. Wilson, and X. Wang. Multi-scale dynamics in a massive online social network. In *IMC*, pages 171–184, 2012.

[31] W.-X. Zhou, D. Sornette, R. a. Hill, and R. I. M. Dunbar. Discrete hierarchical organization of social group sizes. In *Biological sciences*, volume 272, pages 439–44, 2005.