Towards a Characterization of Egocentric Networks in Online Social Networks*

Valerio Arnaboldi, Andrea Passarella, Maurizio Tesconi, Davide Gazzè

IIT-CNR, Via G. Moruzzi 1, 56124 Pisa, Italy
{firstname.lastname}@iit.cnr.it

Abstract. Online Social Networks (OSNs) are more and more establishing as one of the key means to create and enforce social relationships between individuals. While substantial results have been obtained in the anthropology literature describing the properties of human social networks (built "outside" the OSN world), a clear understanding of the properties of social networks built using OSNs is still to be achieved. In this paper we provide a contribution towards this goal, by starting characterizing ego networks formed inside Facebook through the analysis of data obtained from a measurement campaign. Ego networks capture all the social relationships (links) between an ego and all the other people with whom the ego has a social tie. Ego networks are one of the key social structure that have been studied in the anthropology literature, and is thus a reference objective for our work. In this paper we analyze a number of quantitative variables that can be collected on Facebook, which can be used to describe the properties of the social links in ego networks. We also analyze the correlation between these variables and the strength of the social ties, as explicitly ranked by the monitored Facebook users. Our results show that there is an interesting similarity between the properties observed by anthropologists in human social networks, and those of Facebook social networks. Moreover, we found a noticeable correlation between most of the measured variables and the tie strengths, suggesting the possibility of automatically inferring the latter from measurable Facebook variables.

Keywords: Online Social Networks, Measurements, Ego networks

1 Introduction

Social network analysis has become one of the most important multidisciplinary methodology aimed at studying people relationships and information flow within groups of individuals. Social structures are represented as networks of nodes and edges, modelling the set of individuals in a certain context and all the social ties existing between them.

* This work was partially funded by the European Commission under the FIRE SCAMPI (FP7-258414), FET-AWARE RECOGNITION (FP7-257756) and CAPER (FP7-SEC-261712) Projects

Ego networks are a particular kind of social networks representation in which the only relationships considered are those between one person (called ego) and all the individuals directly connected to her (alters). These networks are one of the reference representations of human personal social networks and are largely studied in the anthropology literature [7, 6, 11, 14]. Given two individuals, their tie strength is a numerical representation of the importance of their relationship and is typically used as the weight of the link between them in the social network. The challenging issue in the representation of social networks in general - and of ego networks in particular - is how to find a suitable way to determine the ties strength, given that it should embody all the possible factors underlying social relationships, which can be difficult to be identified.

In [10], the author has given the first informal definition of tie strength, conjecturing the presence of more than one dimension beneath social ties. The author identifies these dimensions as the amount of time, the emotional intensity, the intimacy and the reciprocal services which characterize the relationships. Successive work have added more dimensions to the first definition, achieving a more accurate tie strength characterization. In [2] structural variables such as the number of mutual friends or the dimension of the network are considered as important factors controlling the behaviour of a social relationship. In [11] authors found the frequency of contact to be a good predictor for the social tie strength. Other possible factors controlling for the tie strength are gender, with a direct impact on non face-contact, age and kinship [6].

The diffusion of online social networks (henceforth OSNs), like Facebook, Twitter, LinkedIn and many others, is fostering the availability of a lot of data regarding social interaction between people which was impossible to obtain until few years ago. Several early conjectures made by sociologists on social networks have been confirmed by the analysis of empirical data obtained from the Internet [4, 12, 13]. While ego networks have been well analysed in the anthropology literature, the characterization of ego networks formed in OSNs has received little attention so far [9, 8]. In this paper, we start filling this gap by analyzing the results of a measurement study on Facebook users. The overall goal of this work is starting investigating whether the structure of ego networks formed in OSNs (considering Facebook as representative of them) is similar or not to the structure of ego networks observed in real human networks by anthropologists [7, 6, 11, 14]. To this end, we measured two classes of variables, i.e. *socio-demographic* variables, such as the size of the ego network of the users and various factors possibly impacting on it (e.g., age, gender, ...), and *relational* variables, which relate to the strength of the social relationships between each ego and her alters. In Section 2 we describe in detail our experimental methodology.

In the paper we analyze the distributions of the socio-demographic variables across the users, and of the relational variables across the social relationships of each user. While measurement studies of Facebook users are available in the literature [8, 9], to the best of our knowledge this is the first paper in which the distributions of the quantitative variables related to the ego networks and tie strength dimensions are analyzed. Moreover, the paper also presents an initial

correlation analysis, showing how well each of the measured relational variables correlates with the real tie strength of social relationships. As described in Section 2, real tie strengths have been collected by asking users to explicitly rank the importance of their friends "in their Facebook world".

The key findings of the paper can be summarized as follows (the data and a detailed discussion on their properties are presented in Sections 3 and 4). First of all, we find that the average size of ego networks is in the range of human network sizes found in the anthropology literature [7, 11, 14, 16]. Moreover, the analysis of Facebook variables distributions clearly shows that, from the perspective of an ego, there exist a relative small fraction of "very important" alters, with whom the ego interacts far more than with the rest. As similar features have also been found in human social networks, these results suggest that the structure of OSN ego networks might be quite similar to that of human social networks. Finally, the correlation analysis confirms that the measured variables are sensibly correlated with the real tie strength perceived by the users, and allows us to start understanding which variables, among the measured ones, can better predict tie strengths. This is an initial step towards more refined models, able to automatically estimate tie strengths, only based on the patterns of interactions between users.

2 Data Acquisition

The detailed list of Facebook variables we have identified as best descriptors for ego networks characteristics is listed in table 1. We decided to consider only quantitative variables and to discard all user-filled quantities, which would be prone to the typical problems of semantic analysis.

Relational variables are expected to be related to the tie strength of the corresponding social relationship. Because the tie strength is not directly measurable from Facebook data, this relation is unknown. Thus we decided to collect values of tie strength perceived by the users, asking them to rate their friendships answering the question: *"How do you rate, with a value between 0 and 100, the social relationship between you and this person in Facebook?"*. We collected the answers through an electronic survey. Using a generic question allowed us to capture all the possible definitions of tie strength. With this data we have been able to analyze the correlation between Facebook variables and the social tie strength.

As in [7], we distinguish between active and non-active friends. Active friends are those alters for whom ego spends a non negligible effort to maintain their relative relationships alive. To identify active friends we decided to take into account the set of people who have received a value of tie strength larger than zero. We only considered active friends in our data analysis, leaving the relation between active and non-active networks for future work.

To obtain the data sample for our analysis we performed data acquisition campaign involving 30 people (18 males and 12 females) to whom we asked to use a Facebook application we have built, named Facebook Analyser (FBA).

Table 1: Facebook variables chosen as ego network descriptors

Variable type	**Variable name**
Socio-demographic	age
	gender
	relationship status
	number of friends
	number of active friends
	total number of status updates made
Relational	number of exchanged posts
	number of exchanged private messages
	number of exchanged likes
	number of exchanged comments
	number of mutual friends
	number of tags on the same pictures
	number of days since first communication
	number of days since last communication
	number of tags on the same objects
	number of events attended together
	number of groups in common
	number of likes to the same pages

FBA is able to collect both Facebook variables listed in table 1 and the values of tie strength perceived by the users towards all their friends, with an embedded electronic survey. On average, FBA takes about half an hour for Facebook data acquisition and few minutes for tie strength evaluation. The total duration of data acquisition campaign was three weeks. We have collected a total of 7665 relationships, from which we have extracted 3245 active friendships. While the number of social relationships we have sampled is significant, the number of users involved in the experiment is not sufficient to draw definite conclusions. However, as discussed in section 3, this sample is already sufficient to provide interesting indications on the properties of OSN ego networks, and their similarity with ego networks observed in human social networks.

To study correlations between involved variables, we use the sample correlation coefficient r, which, given two random variables X and Y, is defined as:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{(n-1)S_x S_y} \ , \tag{1}$$

where $\overline{x}$ and $\overline{y}$ are the sample mean, and S_x and S_y are the sample standard deviations, of X and Y, respectively. This is an estimator of the Pearson product-moment correlation coefficient (also known as Pearson's r), defined as:

$$\rho_{X,Y} = \frac{COV(X,Y)}{\sigma_X \sigma_Y} \ , \tag{2}$$

where $COV(X, Y)$ is the covariance and σ_X and σ_Y are the standard deviations. The p values presented with the correlation values regard the correlation significance. Small p-values indicate that X and Y are correlated (see [15] for a precise definition of p).

3 Data Analysis

In this section we present the data obtained from our measurements. A discussion of the main properties that emerge from them is postponed to Section 4.

3.1 Socio-demographic variables

Our experiment involved people randomly chosen within our research area. All the participants were researchers, Ph.D students or master students from 24 to 48 years old ($M = 33.17, SD = 6.78$). One of the aspects we are most interested in is the mean ego-network size. Facebook provides a tool to measure the average number of friends of users, which turns out being around 130. However, this number also considers non-active relationships and unused accounts, thus it can not be used as a reliable measure for our purposes.

The total number of Facebook friends - both active and non-active - of the participants involved in our experiment varies between 59 and 1099 ($M = 255.5$, $SD = 197.59$). However, the active network size ranges between 29 and 368 ($M = 108.17$, $SD = 85.55$). The distribution of this variable, which is depicted in figure 1, is similar to that found by anthropologists [14], [11], studying the properties of human social networks. The mean active network size is also of the same order of that found in human social networks (e.g., 124 in [11]). This number also shows significant correlation with the mean density of the ego networks. ($r = 0.48$, $p < 0.01$). Age does not show a significant correlation with network size. This result is in accordance with [6].

These results provide a first indication of a similarity between the ego networks found in human social networks and those formed by OSNs users.

3.2 Relational variables

To analyse relational variables we split them into more subclasses, to better study the relation between different kinds of interaction and the social tie strength. The classes we have identified are: i) text-based communication variables, ii) like-based communication, iii) homophily variables, iv) time variables and v) structural variables. While most of these classes are self-explanatory, the homophily concept needs a better definition. Homophily measures the social similarity between two individuals. It is the tendency of people with similar social characteristics to have stronger social ties compared to dissimilar individuals.

Text-based communication variables capture all kinds of textual interactions between individuals which can be exploited on Facebook. These variables identify communication styles similar to other traditional methods, such as mail, text

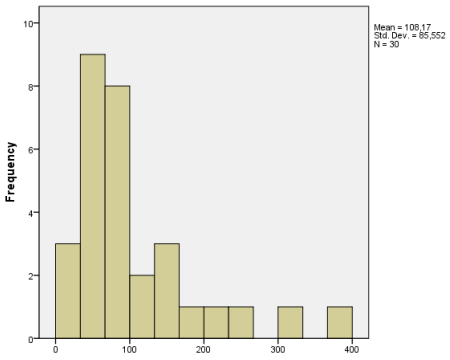
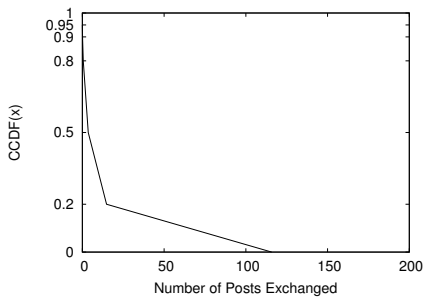
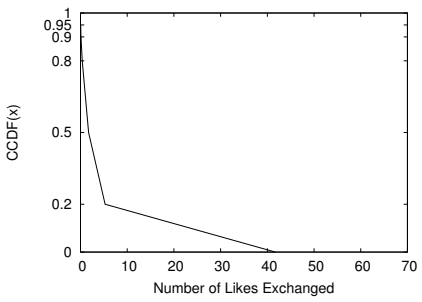


Fig. 1: Active Network Size

messages etc. largely studied by other authors [11] [3]. In these works the evidence of a relation between textual interaction and tie strength has been found. In this class we place variables representing the number of posts, comments and private messages exchanged within Facebook. Figure 2a depicts the CCDF of the number of posts exchanged between ego and alters in our sample. Specifically, we have obtained the indicated percentiles for each user, and averaged them over all users. We don't present the CCDF of the other text-based variables because they show a behaviour similar to posts, which is the most representative variable of the class. Text-based interaction variables show a medium correlation with the perceived tie strength ($r = 0.2$, $p < 0.01$ for comments, $r = 0.2$, $p < 0.01$ for private messages and $r = 0.39$, $p < 0.01$ for posts).



(a) CCDF for the number of posts exchanged between individuals in Facebook



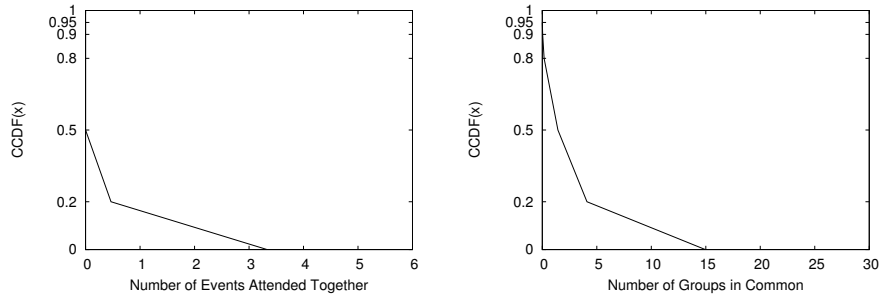
(b) CCDF of the number of likes exchanged between individuals in Facebook

Fig. 2: CCDF of posts and likes

The second class of variables encompasses a novel type of communication, largely used in Facebook, called "like". Likes are special marks left by users on Facebook objects to express a positive feedback. These objects can be posts, comments, pictures, videos and other. The number of likes received from or made

to a certain person could be a good predictor for emotional support, intimacy and frequency of contact of the relationship. Figure 2(b) depicts the CCDF regarding the number of likes exchanged between egos and their friends. As for textual interaction, likes show a long tailed shape. Also like-based communication variables show medium correlation with the tie strength ($r = 0.35$, $p < 0.01$). This result is similar to that previously shown for private messages exchange and displays the high importance of like-based communication inside Facebook.

In the class of homophily variables we have placed the number of groups to which a pair of Facebook friends have a subscription in common, the number of events attained together, the number of likes made on the same objects and the number of pictures in which users appear together. To discover this last quantity we rely on a particular kind of action, typical of the OSNs, called tagging. Tagging is a way to mark a picture with the name of one or more people who appear in it. The presence of two people in the same picture or attending the same event can be seen as good predictor not only for homophily, but also for intimacy and emotional closeness. In Figure 3 we show the CCDF regarding the number of common events and groups (CCDFs of likes on the same objects and tags on the same pictures are qualitatively similar). Homophily variables are more or less equally correlated to tie strength ($r = 0.24$, $p < 0.01$ for events, $r = 0.25$, $p < 0.01$ both for groups and likes).



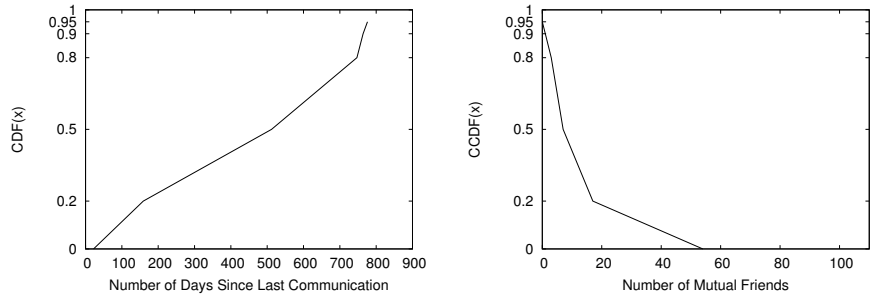
(a) CCDF of the number of events two Facebook friends have attended together

(b) CCDF of the number of groups two Facebook friends have in common

Fig. 3: CCDF of common events and groups

Time variables indicate the frequency of contact and the duration of a relationship. We have considered the time elapsed since the last communication and the time since the first communication, considering all the text-based and like-based interaction variables. Time variables facilitate the distinction between long and short-timed relationships and can discover fading out of social ties - for example if time to first contact and time to last contact are both high presumably the relationship has decayed. Figure 4(a) depicts the CDF of the number of days elapsed since the last communication made between two Facebook users. Time variables are correlated, as expected, with tie strength ($r = 0.23$, $p < 0.01$

for time since first communication and $r = -0.3$, $p < 0.01$ for time since last communication). The negative correlation between time to last contact and tie strength is due to the fact that if the time period elapsed since the last interaction between two people is short, the probability of having a strong tie between them is higher. Time to last contact has been used as the predictor of emotional closeness in most of the work present in the anthropological literature [7, 11, 14].



(a) CDF of the number of days elapsed since the last communication traced between two Facebook friends

(b) CCDF of the number of mutual friends between pairs of Facebook friends

Fig. 4: CDF of the number of days since last communication, and CCDF of mutual friends

The last class of variables contains all the structural dimensions of ego-network ties obtainable from Facebook data. In this class we have identified the number of mutual friends and the age difference between peers as best predictors for tie strength. Age difference between ego and her friends does not show a significant correlation with tie strength. The number of mutual friends is weakly correlated with tie strength ($r = 0.23$, $p < 0.01$). In Figure 4(b) the CCDF of the number of mutual friends is depicted.

4 Discussion

The CCDF of all the relational variables, shown in Figure 2, 3, and 4 present a long-tailed shape. This result suggests the presence of a small set of alters tightly connected to ego and a larger group of people loosely tied to her. Note that this also results for the CDF of the time since last contact, as there is a clear distinction between a set of frequently contacted alters and the rest. These results are similar to the findings in [11] related to human social networks.

The findings in [5] made us intuitively hypothesize the presence of a positive correlation between contact frequency and tie strength. Our results confirm this fact, both in terms of time to last contact and total number of contacts. Comparing the values of correlation seen so far we can say that homophily variables show an unexpected minor influence on tie strength compared to text and like-based

communication variables or time to last contact. Nevertheless homophily could play an important role in the formation of ties rather than in the determination of their weight.

We have found a significant and positive correlation between active network size and network density (i.e. the mean tie strength within the network). This result is in contrast with the findings in the anthropological literature, which have confirmed the tendency of larger network to be less dense [14]. This difference could be explained taking into account that we are studying virtual ego networks, which are only a subset of individuals' complete social networks and users with few friends might not use Facebook to maintain their strongest relationships. Moreover, users with more Facebook friends might spend more time using Facebook than people with few friends, augmenting the weight of the relationships with them. This intuition should be validated with dedicated studies.

As in [6], we have not found a significant correlation between ego age and the size of her network, yet we haven't discovered significant correlation between age and network density (as observed in [6]). This could be due to the low age range of our sample. Authors of [6] have considered a sample composed of participants ranging from 18 to 65 years old ($M = 38.5$, $SD = 13.3$) and they have found significant differences between people placed at the extremes of the sample. Our sample is more homogeneous and does not include those extreme values.

5 Conclusions

In this paper we have presented an initial yet detailed characterization of virtual social ego networks through the analysis of a set of observable Facebook variables. We made a selection over all possible Facebook obtainable information to have a set of variables which at least can contain all the social tie strength dimensions already identified in the sociological and anthropology literature [10, 1, 2], and confirmed by recent empirical studies [9, 8].

We have gathered the values of tie strength perceived by the participants of our experiment towards all their Facebook friends thanks to an electronic survey embedded into our Facebook application. Hence we have analyzed the distributions of the observed variables, and the relation between them and tie strength through correlation analysis.

The results we have found indicate that Facebook active ego networks obtained from our sample data have distribution and mean size similar to those found in other work [14, 11, 16] in the anthropology literature. The distributions of the relational variables we have taken into account have long tailed shapes. This evidence confirms the presence of an high number of weak ties and a lower number of strong ties, as found in [7, 14, 11].

The correlation analysis revealed the relation between the Facebook observable quantities we have considered and the sampled tie strength. Therefore, the relational variables we have identified are good proxies for tie strength. Like-based communication has shown correlation value comparable with all the other types of communication available on Facebook. This is an evidence of the impor-

tance of this style of communication, which is becoming more and more popular in OSNs.

In this work we have not considered the possibility to predict tie strength using the observable quantities we have identified, as done in [9]. As subject of future work, we will target this issue to automatically obtain ego networks representation without asking the users to rate all their friendships. Moreover, we will investigate if specific structures can be identified in OSN ego networks, as it is the case in human social networks.

References

1. Barry, W., Wortley, S.: Different Strokes From Different Folks. American Journal of Sociology 96(3), 558–588 (1990)
2. Burt, R.S.: Structural Holes versus Network Closure as Social Capital, pp. 31–56. No. May 2000 (2001)
3. Culotta, A., Bekkerman, R., Mccallum, A.: Extracting social networks and contact information from email and the Web. In: Collaboration, Electronic messaging, Anti-Abuse and Spam Conference. No. d (2004)
4. Dodds, P.S., Muhamad, R., Watts, D.J.: An experimental study of search in global social networks. Science 301(5634), 827–9 (Aug 2003)
5. Dunbar, R.I.M.: The social brain hypothesis and its implications for social evolution. Annals of human biology 36(5), 562–72 (1998)
6. Dunbar, R.I.M., Roberts, S.: Communication in Social Networks: Effects of Kinship, Network Size and Emotional Closeness. Personal Relationships (2010)
7. Dunbar, R.I.M., Spoors, M.: Social Networks, Support Cliques and Kinship. Human Nature 6(3), 273–290 (1995)
8. Ellison, N.B., Steinfield, C., Lampe, C.: The Benefits of Facebook Friends: Social Capital and College Students Use of Online Social Network Sites. Journal of Computer-Mediated Communication 12(4), 1143–1168 (Jul 2007)
9. Gilbert, E., Karahalios, K.: Predicting tie strength with social media. In: International conference on Human factors in computing systems. ACM Press, New York, New York, USA (2009)
10. Granovetter, M.S.: The Strength of Weak Ties. The American Journal of Sociology 78(6), 1360–1380 (Dec 1973)
11. Hill, R.A., Dunbar, R.I.M.: Social network size in humans. Human Nature 14(1), 53–72 (Mar 2003)
12. Leskovec, J., Horvitz, E.: Planetary-Scale Views on an Instant-Messaging Network. Tech. rep. (2007)
13. Onnela, J., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., Barabási, A.: Structure and tie strengths in mobile communication networks. In: National Academy of Sciences of the United States of America. vol. 104, pp. 7332–7336 (May 2007)
14. Roberts, S.G., Dunbar, R.I., Pollet, T.V., Kuppens, T.: Exploring variation in active network size: Constraints and ego characteristics. Social Networks 31(2), 138–146 (May 2009)
15. Sirkin, R.M.: Statistics for the Social Sciences (2006)
16. Zhou, W.X., Sornette, D., Hill, R.a., Dunbar, R.I.M.: Discrete hierarchical organization of social group sizes. In: Biological sciences. vol. 272, pp. 439–44 (Feb 2005)