Online Social Networks and information diffusion: the role of ego networks

Valerio Arnaboldi[a], Marco Conti[a], Andrea Passarella[a,*], Robin I. M. Dunbar[b,c]

[a]*IIT-CNR, Via G. Moruzzi 1, 56124, Pisa, Italy*
[b]*Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, United Kingdom*
[c]*Department of Information and Computer Science, Aalto University School of Science, Konemiehentie 2, FI-02150, Espoo, Finland*

Abstract

Ego networks models describe the social relationships of an individual (ego) with its social peers (alters). The structural properties of ego networks are known to determine many aspects of the human social behaviour, such as willingness to cooperate and share resources. Due to their importance, we have investigated if Online Social Networks fundamentally change the structures of human ego networks or not. In this paper we provide a comprehensive and concise compilation of the main results we have obtained through this analysis. Specifically, by analysing several datasets in Facebook and Twitter, we have shown that OSN ego networks show the same qualitative and quantitative properties of human ego networks in general, and therefore that, somewhat counter-intuitively, OSNs are just "yet another" social communication means which does not change the fundamental properties of personal social networks. Moreover, in this paper we also survey the main results we have obtained studying the impact of ego network structures on information diffusion in OSN. We show that, by considering the structural properties of ego networks, it is possible to accurately model information diffusion both over individual social links, as well at the entire network level, i.e., it is possible to accurately model information "cascades". Moreover, we have analysed how *trusted* information diffuses in OSN, assuming that the tie strength between nodes (which, in turn, determines the structure of ego networks) is a good proxy to measure the reciprocal trust. Interestingly, we have shown that not using social links over a certain level of trust drastically limits information spread, up to only 3% of the nodes when only very strong ties are used. However, inserting even a single social relationship per ego, at a level of trust below the threshold, can drastically increase information diffusion. Finally, when information diffusion is driven by trust, the average length of shortest paths is more than twice the one obtained when all social links can be used for dissemination. Other analyses in the latter case have highlighted that also in OSN users are separated by about 6 (or less) degrees of separation. Our results show that when we need trustworthy "paths" to communicate in OSN, we are more than twice as far away from each other.

Keywords: Online social networks, ego networks, tie strength, information diffusion, Dunbar's Number

1. Introduction

Online Social Networks (OSNs) such as Facebook and Twitter have introduced radically novel means of interactions among people, which quickly became extremely popular. Complementing more traditional 'offline' means of communication (such as face-to-face communication and phone calls), 'online' social networks are creating a complete virtual social environment, which supports many actions involving social interaction, from extremely simple ones such as "liking" other users' content, up to very complex ones such as looking for a job, advertising products and organizing events. This is one of the most impressive cases of *cyber-physical convergence*, i.e., the process whereby the physical world around us and the virtual world of the Internet are deeply interwoven, constantly interacting with, and dependent upon, each other [1]. In this context, the analysis of the human social behavior in online environments is particularly exciting, as it allows us, on the one hand, to understand features of the human behavior based on huge amounts of data and, on the other hand, to design services and applications exploiting this knowledge. This paper presents a comprehensive overview of a body of work that we have carried out in this field, with the objective of highlighting key structural properties of human personal networks in OSNs, how they are related to known structures of offline human social networks, and how they impact on the patterns of information diffusion that have been observed in these environments.

Undoubtedly, the advent of OSNs also led us to a significant advancement in the study of human online behavior. In fact, big data coming from OSNs represent an invaluable source of information to describe the dynamics of complex social phenomena (e.g., the diffusion of information in the network, the formation of social relationships

*Corresponding author
Email addresses: valerio.arnaboldi@gmail.com (Valerio Arnaboldi), m.conti@iit.cnr.it (Marco Conti), a.passarella@iit.cnr.it (Andrea Passarella), robin.dunbar@psy.ox.ac.uk (Robin I. M. Dunbar)

and communities), which are very difficult to analyze with traditional research methods typically based on surveys and interviews. A lot of effort has been put in the last years to characterize OSNs by studying their graphs, since this is the natural way to study structural properties of human social relationships in OSNs. Most of the literature has focused on *macroscopic* properties of OSN graphs, i.e., the structural characteristics of the global network formed by all users and their connections. A comparatively less explored but equally important subject of investigation are the *microscopic* properties of OSN, and primarily the structural characteristics of our personal social networks, also called *ego networks*. In the anthropology literature it is well known (e.g., [2]) that the characteristics of ego networks are fundamental to determine key facets of human behavior, such as trust, sharing of resources, and formation of communities.

From reference works in psychology and anthropology (e.g., [3, 4, 5, 6]), we know that the properties of offline ego networks are constrained by a series of cognitive and time limits, which bound the amount of relationships that each individual can actively maintain due to their intrinsic cost in terms of 'computational resources' for the brain. Specifically, cognitive constraints limit the total number of active relationships humans can maintain at a non-negligible level of intimacy. This limit is on average around 150 relationships, which is known as the Dunbar's number [3]. The same constraints also dictate specific structures according to which social relationships are organized inside the ego network, as explained in detail in Section 2.

Recent analyses of the structural properties of popular OSNs (i.e., Facebook and Twitter) revealed that online ego networks have the same properties of offline ego networks, with similar size and the same hierarchical structure [7, 8, 9]. This confirms that ego-network properties depend primarily on cognitive constraints of the human brain, and are not influenced by the use of specific communication mechanisms, such as mobile phones [10]and also Online Social Networks. In this regard, Facebook and Twitter do not seem to improve human social capacity, but they simply represent additional social channels we can use. Moreover, in addition to confirming that well-known features of human ego networks also manifest in OSNs, these studies have revealed additional properties [7], which had been hypothesized [11], but never observed due to lack of big data sources. This demonstrates that OSNs can be used also as a 'social microscope' to investigate novel key properties of our social behavior.

Recent results presented, e.g., in [12, 13, 14] show that the patterns of information diffusion we observe in OSNs strongly depend on the structure of the users' ego networks. Therefore, by understanding the latter it is also possible to design OSN services where the features of ego network structures (and the users' behavior they determine) are exploited to optimize data management. For example, as demonstrated by Lerman [15], models of information dissemination in OSNs that consider the constrained nature of human online social behavior overcome some intrinsic limitations of previous state-of-the-art models [16]. Moreover, results presented in [12] have shown that, under the assumption that the ego network structure also defines the level of trust between ego and alters, the length of the shortest path along which trusted information flows between two users in the network can be significantly longer than the few hops (compatible with the well-known 6-degrees of separation concept) previously highlighted in the literature (e.g., in [17]).

In this paper, we first present the main results we have obtained from the analysis of the effect of human cognitive limits on the structural properties of ego networks in OSNs. Then, we present results showing the impact of these ego network structures on macroscopic phenomena, with particular attention to the diffusion of information. In addition, we discuss promising directions to the design of information-centric systems exploiting ego network structures, such as data replication strategies based on ego-network concepts for Distributed Online Social Networks (DOSN), and information dissemination protocols for opportunistic networks based on ego-network cognitive heuristics.

This paper starts from the analysis of human cognitive limits in online environments presented in the book by Arnaboldi et al. [18]. However, while the book focuses on the structural analysis of ego networks in OSNs, this paper focuses on the application of models and analyses based on human cognitive limits to user-centric services and applications, in particular to the diffusion of information.

The paper is organized as follows. In Section 2, we provide the main definitions about online social networks and ego networks, and we present an overview of the most important results of ego network analysis in OSNs. In Section 3, we present the background work in the field of information diffusion modeling, we introduce the most important information diffusion models in OSNs and we discuss their limitations. Then, in Section 4, we present analyses and models of information diffusion in OSNs based on the structural properties of ego networks. In Section 5, we present some research directions where knowledge about social network structures is used to design data-centric services both for OSNs and mobile networks, and we draw the main conclusions of the paper.

2. Ego Network Structural Analysis in Online Social Networks

A typical way of representing social networks is through a graph $G(V, E)$, where V is the set of vertices representing the users and E is the set of edges connecting pairs of users, with each edge representing a dyadic social relationship between the vertices it connects. In OSN graphs, edges can represent, for example, the 'following' relationships of Twitter or 'friendships' of Facebook. Edges may be directed to represent a possible directionality in the

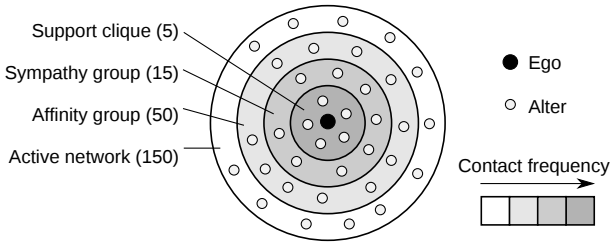

Figure 1: Ego Network Model.

semantic of the social relationship (e.g., in Twitter the difference between a 'follower' and 'followee' relationship). Or, they can be undirected, if dyadic relationships between users are completely symmetric.

The analysis of the OSN graph can be carried out on the unweighted graph, where each relationship is considered of the same type and quality. Typically, the unweighted OSN graph is called the *social* graph. On the other hand, other analyses take into consideration that social relationships are not all equal. To represent this, a weight is associated to each link, and the resulting weighted graph is called the *interaction* graph. For the purpose of the analyses presented in this paper we consider interaction graphs, and we consider that an edge exists between two nodes if the strength of their relationship is greater than zero. To quantify tie strength, we use the frequency of direct contact between users. Considering frequency of contact as a proxy for tie strength (which is a complex concept involving also qualitative aspects) is customary in the literature, and is backed up by many works starting from the classical definition of tie strength given by Granovetter [19]. This relation has later been found in several works on OSNs. See for example the work by Gilbert and Karahalios on the prediction of tie strength by using variables from Facebook [20] and the tie strength predictive models presented by Arnaboldi et al., in which contact frequency in OSNs is the most correlated variable with tie strength [8]. We use tie strength in the interaction graph analysis, as mere 'following' or 'friendship' relationships are not sufficient to define an edge between two users (as it would for the social graph), and at least a minimum frequency of direct communication between them is needed. In the datasets used in the work, a direct communication can be a post on the wall of another user or a comment on a picture for Facebook, or a reply to a tweet created by another user for Twitter.

In the following we consider only the OSN interaction graph. We omit the definitions and analysis of the typical indices used to characterize the macroscopic properties of OSNs and we directly discuss the main microscopic properties found on empirical analyses of OSN graphs. For a discussion on the macroscopic properties of OSNs, we refer the reader to [18].

2.1. Ego-Network Model

Ego networks are one of the key concepts to study the microscopic properties of personal social networks. Different definitions of ego network exist in the literature, corresponding to different approaches in analyzing them. In this paper, an ego network is formed of a single individual (*ego*) and the other users directly connected to it (*alters*) [6]. This model gives particular emphasis to the impact of the ego cognitive constraints on the personal social networks and, in the rest of the text, we will refer to it as the 'Ego-Network Model'. Another possible definition of ego network also considers the links between alters [21], possibly even excluding the links between them and the ego. This is typically used to analyze the topological features of the local social context in which the ego is immersed. Techniques that have been used to this end are based on complex network indices, such as density, connectivity (e.g., Burt's 'Structural Holes' [21]) or ego betweenness [22] measures.

Starting from the Ego-Network model, a fundamental cognitive constraint in the personal social network is the Dunbar's Number [23]. This is the number of relationships that an ego actively maintains in its network over time. The Dunbar's Number in offline ego networks is known to be limited by the cognitive constraints of the human brain and by the limited time that people can spend in socializing. In addition, it is known that cognitive constraints lead people to unevenly distribute the emotional intensity on their relationships. This results in a hierarchical structure of inclusive 'social circles' of alters around the ego (as depicted in Figure 1), with characteristic size and level of tie strength. Specifically, in the reference ego-network model [24], there is an inner circle (called *support clique*) of 5 alters on average, which are considered the best friends of the ego. These alters are contacted at least once a week, and are the people from whom the ego seeks help in case of emotional distress or financial disaster. Then, there is a second layer of 15 alters called *sympathy group* (which includes the support clique) containing close friends of the ego, those contacted at least once a month. After this layer, we find a group of 50 alters called *affinity group* or *band* that contains an extended group of friends. The last circle, called *active network*, contains on average 150 alters (the Dunbar's number) contacted at least once a year. These people represent the social relationships that the ego maintains actively, spending a non-negligible amount of its time and resources interacting with them so as to prevent the corresponding social relationships decaying over time. The sizes of ego network circles form a typical pattern of 5-15-50-150 alters, with a scaling ratio between adjacent circles around 3. This pattern is considered one of the distinctive features of human social networks.

While the focus of this paper is on ego networks in OSNs, we should briefly mention the vast body of work on characterizing ego networks in offline environments. Evidence to support the existence of Dunbar's number and

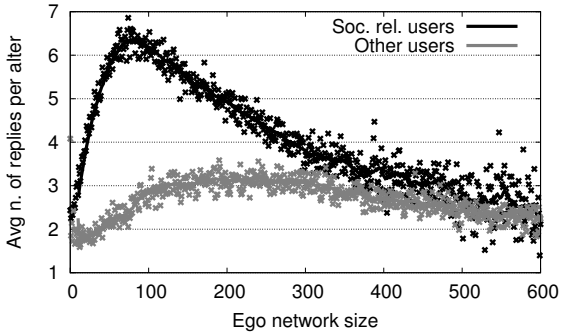

Figure 2: Tie strength as a function of ego network size in Twitter. Points represent the average number of replies made by accounts with different number of friends; thick lines are their running averages.

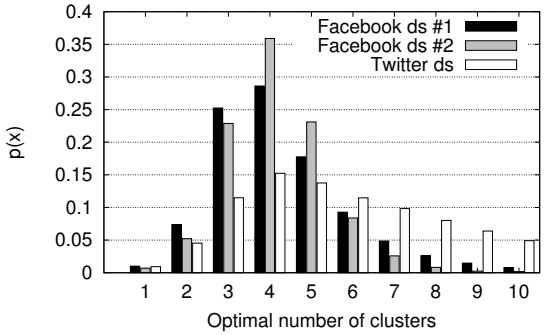

Figure 3: Optimal number of circles for Twitter and Facebook ego networks found through $k-$means and AIC.

the described ego-network structure has come from a number of ethnographic and sociological sources, e.g. [23, 25, 26, 24, 27, 28, 29]. The existence of the circles described in the ego-network model has been also explained as an evolutionary strategy that humans adopted to maintain stability in their increasingly large social groups [30]. More recently, results have been presented on the presence of Dunbar's number [31] and the ego network structure [30] also in phone-call networks. With respect to Dunbar's number, it has been shown that people with a large phone-call ego network spend more time on the phone than people with smaller networks. For phone-call ego network with sizes around 100-150 connections, the total time devoted to phone calls by the egos reaches its maximum. This indicates that beyond this point a very large number of contacts does not imply a proportional increase in the amount of time invested in communication, and this is a clear evidence of time and cognitive constraints in phone-call networks resulting in structures compatible with the general Ego Network Model.

2.2. Ego-Network Structure in OSNs

2.2.1. Dunbar's Number in Online Interactions

A series of analyses conducted on different OSNs have shown that online ego networks have the same structural properties found in offline social environments. Specifically, the work by Gonçalves et al. found the first evidence of the Dunbar's number in Twitter [32]. The authors studied how the average tie strength for the ego networks of Twitter users (calculated as the average number of replies directly sent by a user to its neighbors) changes with the size of the ego networks (the total number of accounts directly contacted with replies by the user). The results have shown that the average tie strength increases with ego network size up to a peak around 100-200, which is compatible with the Dunbar's number, and then decreases substantially. Similarly to the results found in phone-call networks [10], this means that there is a limit on the total amount of social activity also in Online Social Networks, and this has been interpreted as evidence of human cognitive limits that shape Twitter ego networks.

Arnaboldi et al. [33] further refined these results by dividing Twitter users into two classes, the first one containing socially relevant users (i.e., people who use Twitter to communicate with other people and maintain their social relationships) and the second one formed of accounts not directly aimed at the management of personal social relationships (e.g., spammers, bots, companies, public figures). The classification was performed using a supervised learning classifier based on Support Vector Machines (SVM) trained on a set of manually labeled Twitter accounts. The results on the two classes, depicted in Figure 2, highlight, also in this case, a peak at a number of relationships around 100, and show that this peak is a characteristic of socially relevant users only, and is not visible for the other types of accounts.

The peak in the curve in Figure 2 (and in similar curves shown in [32]) highlights the existence of a cognitive constraint limiting Twitter users activity. However, the peak should not necessarily appear at a number of relationships equal to the Dunbar's number, as the latter is defined as the number of relationships that each ego maintains actively in its network, with a contact frequency of at least one message per year (a single message per year in OSNs like Facebook often represents a birthday greeting message, which is the minimum level of interaction to define a meaningful relationship). A series of analyses performed on Twitter and Facebook [34, 8, 9, 7] indicate that the Dunbar's number in OSNs (Facebook and Twitter) is of the same order of magnitude of the values found offline, although being somewhat smaller in some cases (e.g., 90 in Twitter [7]). Finding an OSN Dunbar's number similar to the one in offline human networks results questions the conventional wisdom that OSNs are increasing our capacity to socialize and allow us to maintain an increasing and virtually unbounded number of relationships [35]. The maximum number of active social relationships in online ego networks seems, again, related only to human cognitive constraints, and largely *invariant* with the specific means we use to maintain our social relationships.

There are several reasons behind the presence of somewhat smaller-than-expected ego networks which have been observed in OSNs, both in Facebook and Twitter. As far

Table 1: Size and minimum contact frequency for ego-network circles found through k-means on the contact frequency of Twitter and Facebook users.

Circle	**0**	**1**	**2**	**3**	**4**
			Offline Networks		
Size	?	5	15	50	150
Contact Freq.	?	≥ 48	≥ 12	≥ 2	≥ 1
			Twitter		
Size	1.55±0.02	4.52±0.06	11.17±0.15	28.28±0.32	88.31±0.87
Contact Freq	$\geq$ 276.63±4.06	$\geq$ 113.12±1.49	$\geq$ 49.63±0.66	$\geq$ 16.89±0.21	$\geq$ 2.54±0.02
			Facebook dataset 1		
Size	1.68±0.01	5.28±0.02	14.92±0.06	40.93±0.20	-
Contact Freq	$\geq$ 77.36±0.77	$\geq$ 30.28±0.24	$\geq$ 11.15±0.07	$\geq$ 2.53±0.01	-
			Facebook dataset 2		
Size	1.53±0.03	4.34±0.09	10.72±0.23	26.99±0.61	-
Contact Freq	$\geq$ 58.54±2.62	$\geq$ 22.19±0.74	$\geq$ 7.93±0.23	$\geq$ 1.37±0.04	-

as the Facebook datasets are concerned, early users (the datasets used for these analyses were collected in 2009, when Facebook was still new) may well only have had a small fraction of their offline friends that were present on the platform and typically only sought out people they knew well [7]. In addition, the way datasets were collected (both in Facebook and Twitter) could result in underestimating the number of weak relationships, and this might explain the presence of ego networks smaller than what one might expect from offline analyses [7].

2.2.2. Hierarchical Structure of Ego Networks in OSNs - Number of Circles

To further characterize the structural properties of ego networks in OSNs, a series of studies analyzed the distribution of tie strength within the active networks of Facebook and Twitter users (i.e., considering the people that these users contacted at least once in a year), looking at whether the same hierarchical structure found in offline environments could also be found online. The results have shown that tie strength of online ego networks is unevenly distributed, and relationships can be clustered into circles with level of contact frequency and sizes similar to those found offline. Specifically, Arnaboldi et al. [7] analyzed a large-scale Twitter dataset with approximately 300,000 accounts containing the information about the direct tweets (replies) that each user sent to other social contacts. The analysis of the distribution of the contact frequency of the ego networks has shown that contact frequencies follow a long-tailed distribution within each network, with few strong relationships and many weak ties. This is compatible with the results found offline, and with those found in phone-call communication traces [31]. In addition, to further characterize the distribution of contact frequency for the ego networks, the authors applied a cluster analysis to the frequencies in each ego network, using the $k-$means and DBSCAN algorithms. The rationale of this approach was to seek if clusters of alters could be identified in ego networks, such that contact frequency within clusters is significantly different from contact frequency within the other clusters. If that was the case, clusters would represent the equivalent of layers in the ego network model. To find the optimal number of clusters for each ego network, authors applied $k-$means with increasing values for k and DBSCAN with different possible values for its parameters, and then selected the configuration that yielded the highest value of Akaike Information Criterion (AIC). The goodness of fit obtained for the best configuration has also been measured using the silhouette index. This analysis has been replicated also on two reference Facebook datasets, as explained in [7]. The distribution of the optimal number of clusters in the ego networks of Twitter and the two Facebook datasets, depicted in Figure 3, indicates that social relationships in online ego networks are naturally grouped, on average, into 4-5 circles. This result is compatible with the number of circles found in offline ego networks. It is worth noting that the optimal configurations yield, on average, values of the silhouette index around 0.7, which are high for the type of data analyzed. This indicates that the presence of the circles is not an artifact of the clustering algorithms but rather a real feature of the analyzed data.

2.2.3. Hierarchical Structure of Ego Networks in OSNs - Sizes and Frequencies of the Circles

Following the results obtained for the optimal number of circles, Arnaboldi et al. [7] applied k-means forcing the number of circles to 4 in Facebook and 5 in Twitter for all the ego networks in the datasets, to be able to compare the results in terms of the average size of the circles and their minimum contact frequencies (i.e., the frequencies that define the edges of the circles, in number of messages per year) with those obtained in offline environments. The sizes of the circles, obtained by nesting the clusters found by k-means (remember that each circle is inclusive of all their sub-circles and are thus cumulative sets of clusters obtained by k-means) and the minimum frequencies of contact of the circles are reported in Table 1. For comparison,

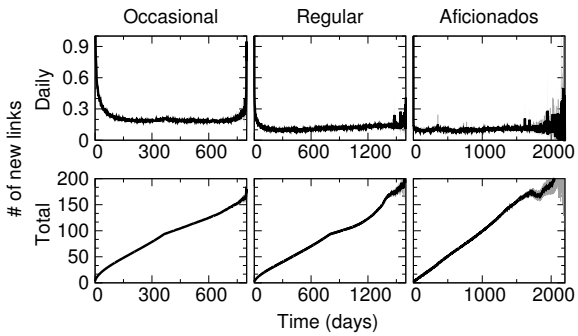

Figure 4: Number of new relationships activated over time in Twitter, on a daily basis and cumulatively.

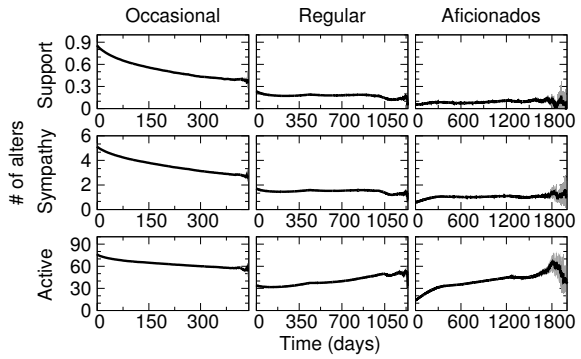

Figure 5: Size of ego network layers over time for Twitter ego networks.

the table also gives the characteristic size and frequencies of the circles found in offline ego networks, as determined by face-to-face contacts [27].

Note that, as explained in the following, the clustering analysis on OSN datasets has consistently highlighted the presence of an additional internal layer inside the support clique, which we denote as layer 0 in Table 1. The contact frequencies of the circles suggest that, in Facebook, alters are contacted approximately at least every five days for layer 0, at least every twelve days for layer 1, at least once a month for layer 2, and at least once every six months for layer 3. These values are compatible with those obtained offline. In the Twitter dataset, the contact frequencies are higher – i.e., at least once every one/two days in layer 0, at least every three days in layer 1, at least once a week in layer 2, at least once a month in layer 3, and at least two/three times a year in layer 4. This can be attributed to the specificity of Twitter, which is explicitly designed for the exchange of short and frequent messages between users. Bearing in mind this difference between Facebook and Twitter, Arnaboldi et al. [7] matched (as reported in Table 1) the layers found in online ego networks with those in face-to-face networks. This allowed them to conclude that the hierarchical structure of ego networks in OSNs is similar to the structure found in offline ego networks. In addition, the new inner circle (Circle 0), with an average size of 1.5 inside the support clique (thus named "super support clique") fits perfectly in the hierarchy of ego network circles, presenting a scaling ratio for the size of around 3 with respect to the next circle (i.e., the support clique). The existence of this structural element of ego networks was hypothesized since long in the anthropology community [11]. However, before [7], no large enough dataset was available to identify, with sufficient statistical confidence, a layer composed of such a small number of social relationships. To the best of the authors knowledge, therefore, results presented in [7] are the first empirical evidence confirming this hypothesis.

2.2.4. Evolution of Online Ego-Network Structural Properties Over Time

The analyses presented so far evinced that the structural properties of online ego networks are the product of a series of time and cognitive constraints on the social capacity of the users, and that these properties are similar to those found in offline environments. However, these analyses built ego networks from a 'static view' of the communication activity of the users, and do not account for possible variations over time of the analyzed structures. In this section, we present a series of studies on how the structural properties of ego networks evolve over time. These analyses further describe the strategies adopted by people to cope with their limited resources for the management of social relationships over time.

.

Number of New Relationships Activated Over Time. As far as complete ego networks and their evolution over time are concerned, several analyses showed that users add new social relationships in their ego networks at a higher rate when they join the platform, and then they tend to maintain a constant growth rate over time. This has been found in phone-call networks [36], as well as in OSNs [37, 38, 39], and can be seen in Figure 4, which depicts the average number of new alters added by egos in Twitter over time, both on a daily basis and cumulatively [38, 18]. This means that egos constantly add new relationships in their ego networks, and thus the composition of ego networks constantly changes over time.

Activation and Deactivation of Relationships - Turnover Process. Perhaps even more interestingly than the fact that ego networks show a constant rate of activation of new relationships over time, several analyses showed that the total number of active relationships in the ego networks remains constant over time, thus showing a balance between the number of relationships that are activated/deactivated [38, 36]. This balance creates a 'turnover' process – a strategy adopted by the egos to cope with their limited social resources. It is worth noting that also the

Table 2: Average Jaccard coefficient of different network layers

layer	Occasional	Regular	Aficionados
active net	0.191	0.190	0.193
sympathy gr.	0.287	0.309	0.362
support cl.	0.346	0.395	0.488

size of each ego network circle remains quite constant over time (as can be seen in Figure 5, and further described in [38, 18])[1].

As a measure of the turnover in the Twitter ego networks analyzed in [38], the authors calculated the Jaccard coefficient between snapshots of the ego networks (divided into the different layers) of one year each. To analyse turnover at each layer, ego networks that always maintain a non-empty support clique in all the one-year windows were analysed. The results are reported in Table 2.

We note that the turnover is in general quite high, always higher than 51.2%. It is about 81% (Jaccard coefficient ~ 0.19) for the entire ego network. The sympathy group shows a percentage of turnover between 71.3% and 63.8%, whereas the support clique is between 65.4% and 51.2%. These results denote a behavior similar to ego networks in offline social networks, where the inner layers contain stronger relationships that should be intuitively less affected by the turnover in the network.

From the results presented so far, it is clear that the analysis of online ego networks reveal the presence of human cognitive limits that shape *in a similar way* the structure of personal social networks across many social interaction means, from online platforms to face-to-face interactions. In the next sections we survey several results showing that these cognitive limits have a strong effect on global social processes, and specifically on the diffusion of information in online networks. With respect to this aspect, before delving into the details of information diffusion analyses , we give a brief but exhaustive introduction to the most widely adopted information diffusion models in OSN analysis.

3. Information diffusion models in OSNs

The study of the process underpinning the diffusion of information in OSNs is one of the most important research aspects in the field of social network analysis. Following the categorization given in [40], there are three main research tasks related to information diffusion in OSNs: (i) detecting popular contents, which have a high probability of spreading to a large number of users; (ii) modeling the diffusion of information by identifying the paths that it is likely to follow in the network (i.e., tree-shaped paths called 'cascades' of adoptions); (iii) identifying influential spreaders, which are nodes from which large cascades can be generated. In this paper, we focus on the second and third aspects of information diffusion in OSNs, as they are directly influenced by the presence of human cognitive limits, whereas the first aspect is more related to the type of contents circulating in the network, and it should not be influenced by these constraints, at least not directly.

Information diffusion models are generally divided into *explanatory* models and *predictive* models. The former models start from the observation of real information cascade traces collected from OSNs and they aim to find a set of parameters that maximize the likelihood of the observed data. On the other hand, predictive models try to reproduce human behavior in the diffusion process by defining a set of 'rules' that each single node should follow when exposed to information, to decide whether to diffuse it or not. Clearly, the design of predictive models generally requires a more detailed knowledge of the properties of human social behavior and it is more directly exposed to the presence of cognitive constraints than explanatory models. For this reason, in this paper we will focus on predictive models. The interested readers can refer to [40] for an exhaustive discussion of explanatory models.

Predictive models can be further categorized into two groups: (i) non graph-based approaches [41, 42, 43], and (ii) graph-based approaches [44, 45, 46]. The first type of model does not assume the existence of a specific graph structure and borrows its main concepts from epidemiology. Non graph-based models split nodes into different classes (i.e., different states in which the nodes can be during the diffusion process). For example, a node that has obtained the information and is going to share it further is placed in the class of 'infected' nodes, whereas the other nodes can be 'susceptible' if they are not infected but can be infected in the future, or 'recovered' if they were infected in the past, but they "recovered" from the infection and they cannot be infected anymore. The different models define a set of rates of transition between the states for the nodes. For example, SIS and SIR define the possible transition for the nodes from 'susceptible' to 'infected' and again to 'susceptible' (for SIS) or to 'recovered' (SIR). There exist more refined versions of these models specifically designed for OSNs, but we will not delve into their properties in this paper, because they are known to suffer from severe limitations , as they fail to consider the influence that social relationships existing between nodes has on the diffusion. In fact, in OSNs, differently from epidemiological processes, the diffusion is highly influenced by the existence of relationships between people [47, 48], and indeed the interplay of human cognitive limits and network structure differentiates the spread of information from other social contagions [15].

Graph-based models specifically start from the assumption that a node is willing to fetch and further share a content (i.e., it is infected) if it is exposed to it from one (or

[1]Note that the particularly small sizes for the inner layers in Figure 5 (between 0 and 1 for the support clique) are due to the presence of some ego networks without inner layers. This is a by-product of the methodology used in [38] for the identification of ego-network circles, which, for computational reasons, is based on pre-defined levels of contact frequency and not on cluster analysis.

Table 3: Information diffusion properties of ego network rings in Twitter, where x and y are $\overline{ts}$ and $\overline{diff}$. r_{xy} is the Pearson's correlation between x and y and $\hat{\beta}$ and $\hat{\alpha}$ are respectively the estimated intercept and angular coefficient of a linear model that relate tie strength and the diffusion, fitted through linear regression. Each ring is the exclusive part of each ego network circle that is not included in any internal circle.

	all alters			**human alters**			**other alters**		
Ring	r_{xy}	$\hat{\beta}$	$\hat{\alpha}$	r_{xy}	$\hat{\beta}$	$\hat{\alpha}$	r_{xy}	$\hat{\beta}$	$\hat{\alpha}$
R_1	0.61	0.49	0.03	0.80	0.74	0.03	0.74	0.58	-0.01
R_2	0.52	0.62	0.01	0.76	0.76	0.02	0.71	0.59	0.02
R_3	0.44	0.74	0.00	0.72	0.80	0.03	0.67	0.64	0.02
R_4	0.34	0.97	0.00	0.66	0.85	0.06	0.65	0.72	0.02
R_5	0.22	1.58	0.00	0.61	0.99	0.09	0.65	0.93	0.03
Whole net (C5)	0.46	0.57	0.02	0.68	0.83	0.09	0.65	0.78	0.03

more) of its social contacts, and the probability that the node will be infected is proportional to the importance of the relationships with the infected neighbors. The simpler models in this class are the *Independent Cascades* (IC) and the *Linear Threshold* (LT) models. Both IC and LT proceed at discrete time steps. In IC, at each step n, the nodes that have been infected at the previous time step (or the initial spreaders – called seeds – if the time step is the first one) infect each of their neighbors with a probability that is proportional to the weight of their links. In LT, each node is infected at a given time step if the sum of the weights of its incident edges connected to already infected neighbors is above a given threshold. IC and LT diffusion processes stop when no new nodes are infected during a certain time step. These models, despite being simple, have been largely used to model information diffusion in OSNs. The weights on the edges of the network graph, which define the probability that the diffusion will pass through the links, are generally fixed (equal for all edges), or, in other cases, they are derived with maximum likelihood estimation from the observations of real diffusion cascades [49].

As recently discussed by Lerman [15], the diffusion cascades generated by IC and LT models are far from being realistic, as they often largely differ from real cascades originated within OSNs. In particular, modeled cascades often reach all the nodes in the network, whereas large diffusions are extremely rare in reality. As demonstrated in [15], including parameters based on human cognitive limits in the diffusion models is sufficient for achieving higher accuracy, and to better reproduce human social behavior. This author, for example, shows that models which include features regarding the position of a post received by a user in its Twitter timeline and the popularity of the post increase the accuracy of the prediction, as people have limited cognitive resources to spend for the diffusion process and they generally focus on fresh and popular contents.

Despite this first attempt to improve information diffusion models using concepts related to human cognitive limits, there is still large room for improvement. Since the structural properties of ego networks, as we have seen in Section 2, are directly related to the presence of human cognitive limits and they can be directly measured from OSN data, their analysis can help to identify additional features to improve the existing information diffusion models.

4. Combining Ego-Network Structural Properties and Information Diffusion Models

In this section, we present the work we have done to improve the accuracy of OSN information diffusion models through information on the structure of ego networks. Firstly, we present an analysis of single-hop information diffusion in Twitter and its relation with tie strength between users. Then, we focus on individual nodes, and analyse how ego-network features of a node are correlated with the information cascades originating from it. Finally, we consider information diffusion at the entire network level: Assuming that the tie strength is correlated with the level of trust between egos and alters, we analyse how trust between users impact information diffusion at the level of the entire OSN.

4.1. Analysis of Information Diffusion at the Ego Network Level – Single Hop Diffusion

One fundamental reference to understand information diffusion in human social networks, and how it is correlated with tie strength, is the seminal work by Mark Granovetter [19]. Granovetter hypothesised that weak ties, although being less frequently activated than strong ties, provide access to diverse information in the network, speeding up the diffusion process. In addition, the larger number of weak ties with respect to strong ones in ego networks makes the cumulative quantity of information that passes through them exceed that circulating through strong ties. It is also worth noting that the presence of too many strong ties might 'trap' information in cliques of users and could slow down the diffusion. From these observations Granovetter argued, and proved experimentally, that weak ties are fundamental for information diffusion, coining the well-known expression "the strength of weak ties".

Table 4: Correlation analysis between nodes and cascades' properties

	Node Coverage					**Cascade Depth**				
	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	$\alpha = 0.5$
Unweighted Social Graph										
Degree	0.15	0.14	0.17	0.20	0.23	0.26	0.25	0.27	0.28	0.29
Clust. Coef.	0.05	0.03	0.01	−0.02	−0.05	−0.05	−0.05	−0.07	−0.08	−0.11
PageRank	−0.09	−0.08	−0.06	−0.05	−0.03	−0.13	−0.11	−0.10	−0.08	−0.07
Eigenv. Cent.	0.32	0.27	0.29	0.30	0.29	0.35	0.34	0.34	0.34	0.33
Weighted Social Graph										
Activity	0.68	0.71	0.77	0.83	0.87	0.72	0.75	0.77	0.78	0.79
Clust. Coef.	0.16	0.14	0.11	0.09	0.06	0.09	0.10	0.08	0.06	0.04
PageRank	0.27	0.31	0.37	0.43	0.49	0.32	0.34	0.36	0.39	0.41
Eigenv. Cent.	0.48	0.57	0.55	0.54	0.51	0.20	0.28	0.29	0.30	0.30
Burt Constr.	−0.20	−0.18	−0.21	−0.22	−0.25	−0.35	−0.34	−0.36	−0.37	−0.38

Starting from the hypotheses of Granovetter, several research studies confirmed the importance of weak ties, but also highlighted that strong ties typically "carry" a very significant flow of information between egos and alters. Specifically, it has been shown that there is a positive correlation between the tie strength of a OSN social relationship and the amount of information diffused over that relationship. This has been observed, for example, in the diffusion of posts containing URLs in Facebook [47], where tie strength has been measured as the frequency of direct Facebook posts between users. In Twitter, the work recently presented in [13], based on the same dataset of the work presented in Section 2, showed that there is a significant positive correlation between the volume of retweets on Twitter social relationships (i.e., the volume of information diffused through the link) and the frequency of direct contact (i.e., the frequency of Twitter replies exchanged between the involved users – a measure of tie strength). Specifically, the authors considered the tie strength as the contact frequency between users normalized by the total contact frequency of each user. This ensures a homogeneous analysis, eliminating differences between ego networks due to their different duration or the different frequency of Twitter use of the users. As a measure of information diffusion, the authors used the volume of retweets passing through a social link. Therefore, the tie strength between an ego e and one of its alters a has been calculated as the percentage of frequency of replies sent by e to a with respect to the total frequency of replies of e. This measure is expressed by the following equation:

$$\overline{ts}_{e,a} = \frac{\text{reply frequency from } e \text{ to } a}{\text{total reply frequency of } e}. \quad (1)$$

Similarly, given a link between the ego e and alter a, the frequency of retweets done by e of a's tweets has been normalized by the total retweet frequency of e, as follows:

$$\overline{diff}_{e,a} = \frac{\text{link retweet frequency}}{\text{ego total retweet frequency}}. \quad (2)$$

The results of the analysis, reported in Table 3, indicate that the correlation between $\overline{ts}_{e,a}$ and $\overline{diff}_{e,a}$ has medium/high values, and it increases from the outer to the inner parts of ego networks (from weak to strong ties), with values greater than 0.6 for the innermost layer. This confirms the significant level of positive correlation between tie strength and information diffusion at the level of individual social relationships. Very interestingly, when alters in each ego network are divided into human and non-human users (according to the same classification explained in Section 2), correlations are significantly higher (close to 0.8 for human alters in the innermost layer and always higher than 0.6 for the other layers). This indicates a significantly different diffusion process for human and non-human alters.

To better investigate the possibility of predicting the rate of information diffusion on social relationships from tie strength, the authors also performed a regression analysis on the two measures, by studying the relation between tie strength and information diffusion variables, expressed by the following equation.

$$\overline{diff} = \alpha + \beta * \overline{ts} \quad (3)$$

The estimated parameters found through linear regression for the equation are reported in Table 3. It is worth noting that the values of β, when human and non-human alters are analyzed separately, increase from inner to outer layers. This means that, although tie strength decreases when moving from inner to outer layers, the diffusion rate does not decrease at the same pace. This confirms the importance of weak ties for information diffusion: all in all, a strong tie carries a higher flow of information than a weak tie (this is shown by the positive correlation and by the high values of β also in inner layers). However, the rate of diffusion "per unit of tie strength", i.e., β (since α is always close to 0), is higher for weak ties. This is a strong indication that information over weak ties is very precious for the ego, and diffuses less dependently on the level of tie strength. Moreover, this also explains the lower values of correlation between tie strength and the diffusion rate

Table 5: Percentage of nodes of the original graph covered by the largest component for the different graphs created by considering only the specified circles of the ego networks of the users.

Ego Network Circle remaining in the network	**Percentage of nodes (and links) in the original graph present in the reduced graphs**					
	No reinsertion	Highest Frequency	Lowest Frequency	Probabilistic	Inverse Probabilistic	Random
Active network	0.966 (0.219)	0.994 (0.222)	0.994 (0.222)	0.994 (0.222)	0.994 (0.222)	0.994 (0.222)
Affinity group	0.297 (0.046)	0.714 (0.094)	0.705 (0.093)	0.726 (0.095)	0.722 (0.095)	0.725 (0.095)
Sympathy Group	0.191 (0.028)	0.642 (0.081)	0.634 (0.079)	0.661 (0.082)	0.657 (0.081)	0.661 (0.082)
Support Clique	0.028 (0.004)	0.386 (0.065)	0.385 (0.063)	0.453 (0.066)	0.444 (0.065)	0.456 (0.065)

in outer layers. Note that, when human and non-human alters are mixed together, this process is less visible.

4.2. Diffusion models based on Ego Network Structure – Multi-Hop Diffusion

To extend the results previously presented to complete information diffusion models, Arnaboldi et al. [50] introduced a novel information diffusion model based on the basic mechanism of the IC model, but with features specifically defined to consider human cognitive limits on the behavior of single nodes . The model is built starting from an OSN graph extracted from a large-scale Facebook communication dataset (Facebook dataset 1), and the probability of diffusion on each relationship is calculated as the frequency of contact between the involved users, normalized with respect to the maximum value in the network, multiplied by an aging factor equal to $(1-\alpha)^{n-1}$, where n is the time step of the diffusion process (as in the original IC model – see Section 3), and α controls the speed of aging of information. The model considers, on the one hand, that the diffusion rate is proportional to tie strength on the edges between users, which is directly derived from the communication traces in the dataset. In this way, the probability of diffusion for the relationships of each ego network follows the same distribution of the tie strength in the ego network, which, as we have seen in the previous section, is shaped by cognitive constraints. On the other hand, the model considers higher probabilities of diffusion for fresh information, and all the probabilities are decreased exponentially as time passes and information gets old. This behavior is in line with the idea that the cognitive constraints of the human brain make users prefer to diffuse fresh information and discard older messages [15]. Information cascades generated by the proposed model are more similar to real cascades (i.e., they show a similar distribution of depth, with long-tailed shape and a very low probability to produce extremely large cascades) than those generated without considering the aging factor (i.e., setting $\alpha = 0$, and thus using the standard IC model) [50].

Since the diffusion cascades generated by the model presented in [50] are generally short, in particular for values of α greater than 0.1, the authors performed also a detailed correlation analysis between several centrality indices of the nodes from which these limited diffusion starts and the size and depth of the resulting diffusion cascade trees. As cascades are short, intuitively the length of the cascades might significantly depend on the local structure of the ego networks of their seeds. The results, reported in Table 4, show that classical centrality indices of the unweighted network graph (e.g., node's degree, local clustering coefficient, PageRank – considering only the existence of social relationships in the network and not their weight) have very low values of correlation with the size and depth of cascades, but when these indices are calculated on the weighted network graph, thus taking tie strength into account, the correlation values are much higher and sufficient to identify influential spreaders (at least for the short diffusions analyzed in the aforementioned work) only using the properties of their ego networks. It is worth noting that the network index with the highest correlation values is the total activity of the user with respect to its ego network, i.e., the total contact frequency of the ego, calculated as the sum of the frequencies on its links.

4.3. Impact of Trusted Relationships and Ego Network Layers on Information Diffusion in Complete OSNs

4.3.1. Network Coverage

Another important aspect we have analyzed on the interplay between ego networks and information diffusion is the impact of each single layer of the ego network on the diffusion of information over the entire OSN. In this case, the basic assumption is that tie strength is positively correlated with trust between the users having a social relationship. Therefore, layers in an ego network can be seen as a way to group alters with a similar trust level for the ego. Exploiting this concept, the authors of [12] performed a study on a large-scale Facebook graph (Facebook dataset 1), where they incrementally removed edges from the network according to their membership with respect to the circles of the ego network of each user. For example, as a first step of the analysis, they removed all the relationships outside the active network of each user, keeping in the graph only the relationships with a contact frequency higher than one contact per year. Then, the authors studied the structural properties of the resulting global network graph to see whether this had the same properties of the original graph in terms of its information diffusion capacity. Specifically, they considered the percentage of nodes that remain in the giant component of the network, and are thus reachable by information propagating in such com-

Table 6: Average shortest path length of graphs created by considering only the specified circles of the ego networks of the users with respect to the original Facebook graph.

	Ego Network Circle remaining in the network			
Re-insertion Strategy	*Active network*	*Affinity group*	*Sympathy Group*	*Support Clique*
No Reinsertion	11.67	10.81	10.51	11.07
Highest Frequency	11.72	11.75	11.95	13.74
Lowest Frequency	11.68	11.93	12.19	16.11
Probabilistic	11.71	11.95	12.21	16.16
Inverse Probabilistic	11.71	11.97	12.30	17.42
Random	11.74	11.95	12.28	17.15

ponent, with respect to the total number of nodes in the original graph. Note that this analysis is equivalent to hypothesizing that relationships above a certain level of trust (identified by the tie strength of the selected ego network circle) diffuse information with probability 1, and those below this threshold never diffuse information. If, after filtering out relationships that are below the selected threshold, a large number of nodes with respect to the original graph remain inside the giant component of the network, they will eventually be reached by information. On the other hand, nodes which end up disconnected from the giant component are not reachable by information. The results, reported in Table 5 under the column "No reinsertion", show that the graph obtained by removing edges outside the active networks of the users has a very high percentage of nodes inside its giant component with respect to the number of nodes in the original graph (approximately 97% of the original network - against a deletion of about 78% of the links in the original network). However, when the next layer of the ego networks is removed, the size of the giant component in the resulting graph drops to a value under 30% of the size of the original graph. Note that this is obtained after removing only an additional 15% of links. This means that limiting the diffusion to the active network of the users does not excessively restrict the diffusion, even though a lot of links are not used anymore. But a comparatively much milder reduction of weak ties inside the active networks of the users significantly limits the diffusion capacity of the network. This indicates that, while very weak ties are not too important for supporting information diffusion, those within the active circle of the users are fundamental for making all nodes reachable from each other. Finally, removing additional layers further limits information diffusion, reaching less than 3% of the network when only alters in the support clique are used.

Arnaboldi et al. also proposed a strategy for improving network connectivity for the graphs obtained by using only relationships at a certain level of trust (i.e., inside a given ego network circle). The rationale of the study was as follows. Limiting diffusion at a certain layer means that information propagation can occur only over links with a certain level of trust. However, based on the results in Table 5, this may result in very limited diffusion. Therefore, it is interesting to understand whether a higher diffusion can be achieved only by including *all* relationships with lower levels of trust, or it is sufficient to include only *a few* of them and, in this case, using which criterion. The results suggest that even re-inserting in the network graph a single relationship for each ego network is sufficient for obtaining significantly larger connected components, even when only the innermost circles are kept from the original graph. They tested several possible strategies to choose the relationship to re-insert in each ego network, from a completely random choice ("Random" in Table 5) to taking the relationship with higher (or lower) contact frequency ("Highest Frequency" and "Lowest Frequency"), or according to a probability proportional (or inversely proportional) to the contact frequency of the relationship ("Probabilistic" and "Inverse Probabilistic"). The strategies that provided the greatest increase of information diffusion, as reported in Table 5, are the "Probabilistic" and the "Random" ones. Considering the cost for the users, i.e., the risk it takes by including less trustworthy alters in the diffusion process, "Probabilistic" is better than "Random" since it guarantees that, on average, the re-inserted nodes have higher trust level than randomly selected nodes.

4.3.2. Average Path Length

The analysis presented in [12] considered also a different possible aspect of the network graphs (obtained by deleting relationships outside a specific ego-network circle) to quantify their capacity of diffusing information. Specifically, the authors took into account the average weighted path length of the giant components of the resulting graphs as a measure of how easily information can circulate inside them. To this end, the inverse of tie strength is considered as the cost associated to each link, or, in other words, the cost for a user to share a message over that link. The cost of a path is therefore a measure of the "lack of trust" or of the amount of resources to be spent to guarantee trusted communications between users, considering the type of social relationship between them. For each pair of nodes in the network, the best path between the nodes is selected as the path with lowest total cost with respect to all the other possible paths. The average path lengths (thus the number of relationships in the least cost paths) for the different graphs are reported in Table 6. For comparison,

the average path length of the original unweighted graph is around 5, as reported, e.g., in [51, 17]. Moreover, the average path length on the *entire* weighted graph (i.e., including weak ties outside of the active networks) is about 10. When removing parts of the active network, the average path length only slightly increases with respect to the entire weighted networks. Specifically, it is always approximately equal to the one over the unweighted graph. Interestingly, the length of the shortest path does not depend much on the set of layers excluded from the diffusion process, or on the re-insertion policy. Note, however, that when only more internal layers are used, the coverage is significantly lower (see Table 5). Thus, although information "travels" approximately the same number of hops, it remains "trapped" close to where it originated.

One of the most interesting results from Table 6 is that the path length is much higher than what one would expect on the unweighted graph, showing that, when considering trusted paths, users are significantly farther away from each other than the well-known anecdotal six degrees of separation, a conclusion that has been confirmed by recent analyses of the Facebook *unweighted* graph [17]. This is because, when one considers a cost associated with information propagation, which is determined by the strength of social ties, weak ties become much less used than strong ties. Therefore, instead of using weak ties to bridge across distant areas of the social network, information is diffused through a higher number of stronger relationships (through which there is a higher probability of diffusion), which however are not able to propagate information across the network as quickly as weak ties would do.

5. Conclusion and Future Research Directions

In this paper we have presented our most recent work on the characterization of the structural properties of social relationships in OSNs, and how they depend on human cognitive and time constraints. From our analyses, we have seen that the properties of ego networks in OSNs are compatible with those found in offline environments. This indicates that the hierarchical structure of concentric layers of alters around the ego is consistent among different social environments, and is not influenced by the use of a particular communication medium. This is a clear indication that human cognitive and time constraints shape social relationships not only in offline environments, but also in OSNs, in contrast to the conventional wisdom that OSNs are able to improve our social capacity and allow us to maintain a much larger number of relationships than is possible "offline".

In addition, we have shown that the structural properties of online ego networks can be used to understand in detail the process of information diffusion in OSNs, and to create more accurate predictive diffusion models. Tie strength is highly correlated with the amount of information that flows through each link, even though the correlation is higher for inner layers. This is consistent with the well-known Granovetter's results, that showed that strong ties can carry a significant amount of information, although weak ties are also important for acquiring diversity of information (confirmed in our findings by the comparatively lower correlation between tie strength and amount of information). Moreover, properties of ego networks are highly correlated with the depth and size of information cascades originated from the ego. Finally, we have assessed the impact of tie strength on the diffusion of trusted information, showing that the well-known result about six (or fewer) degrees of separation does not hold when information can flow only over social relationships above a certain level of trust.

The relevance of ego-network structure in the study of OSN properties, opens several research directions. An aspect that has not been discussed in this paper, but that can be important for further improving information diffusion models, is the creation of generative models of social networks (i.e., models able to generate synthetic social network graphs) with properties similar to those of real social networks. Adding features related to the structural properties of ego networks in the models can lead to social network graphs with structures that reproduce those found in real (online) social networks in a more accurate way. For example, [52, 53] propose a new generative model of social network graphs able to create a synthetic weighted network with a set of microscopic (ego network) and macroscopic (complete network) properties given as input. According to the model, an ego network for each user is built iteratively following a set of distributions for the sizes of the ego networks and of their circles, and for tie strength. While they are being generated, ego networks are also combined together, forming a complete social network graph. To do so, each ego is associated with an agent, that, at discrete steps, adds a new alter into its ego network, placing it in one of its circles according to the defined distributions. Each agent stops when its ego network reaches the size that has been assigned to it. At each step, the agents that have not yet completed their ego network select a new node to connect to. As reported in [52, 53], this model is able to reproduce both macroscopic and microscopic properties of reference networks on which it has been validated. In particular, as demonstrated through a detailed validation performed on a large-scale Facebook dataset (Facebook dataset 1), the model preserves the nodes' degree distribution, the average shortest path length, and the clustering coefficient of the reference networks. The graph generated by the model also preserves the fundamental properties of the ego network model and the size and tie strength distribution of the layers compatible with those of the reference network. The synthetic social network graphs generated by this method are very versatile tools to analyze *in silico* the information diffusion process in social network graphs with different possible structures, and to find a possible relation between these structures and the intrinsic capacity of the network to diffuse contents.

The dependence of information diffusion on the trust

of social relationships discussed in Section 4 can have also a significant impact on the design of novel social networking platforms such as Distributed Online Social Networks (DOSN), as discussed in [12]. Examples of DOSN include Diaspora [54], Peerson [55] and Safebook [56]. DOSN implement the functionalities of OSN platforms, but in a completely decentralized way. In fact, personal data of the users and the content they exchange is stored directly on their devices, without the need of any third party server to operate the social networking platform. This provides much more control to the user over their personal information, but requires caching and replication techniques to guarantee data availability. In fact, nodes can suffer disconnections from the network or may be switched off for long periods. A typical solution to achieve data availability in DOSN is to replicate data on 'trusted' peers – i.e., each user gives a copy of its data to one or more nodes in the network with which it has a level of trust higher than a certain threshold, so as to avoid sending personal information to untrustworthy and potentially fraudulent nodes. The results of the analysis previously presented indicate that, with an appropriate design, by choosing these nodes within the active network of each user, and even by using only some of the ego network layers (plus a few additional selected relationships outside them), does not significantly prevent information from being reachable from any other parts of the network. This is the concept at the basis of the work presented in [57], where the authors implemented a DOSN with a replication strategy based on the structural properties of the ego networks of the users. Nodes chosen for hosting data replicas can change dynamically according to users' churn, but are always picked among the active network of each user. The results of the work showed that with a maximum of 2 replicas for each user with at least 40 social relationships, data availability is always higher than 90%.

As another promising future research direction, the structural ego network indices presented in this paper can be used to improve data availability also in other types of social-oriented networking systems, such as Mobile Social Networks (MSN). In MSN, users directly generate and share contents with nearby users in real time by exploiting the physical interactions of their personal mobile devices such as smartphones by exploiting opportunistic networking techniques [58]. Knowledge about the structural properties of ego networks could both improve the accuracy of data dissemination in MSN, and can make it more easily adaptable to different social contexts. For example, the exchange of information between users in proximity through opportunistic networks can be optimized by relying on social circle cognitive heuristics applied to information diffusion policies [59]. Social circle heuristics are models of human cognitive functions developed in the cognitive psychology research community, to describe the mental mechanisms that induce an individual to acquire information as an effect of the availability of the same information on its social neighbors. Relationships in the 'social circles' of each individual have different influence on its information acquisition actions. In [59], the authors implement a completely decentralized and self-organizing algorithm whereby nodes, upon encountering with each other, decide which information to exchange based on a decision process determined by the social circle cognitive heuristics. Performance results show that this approach, compared to algorithms that do not exploit knowledge about the social structures of users relationships, is able to obtain a similar level of efficiency in terms of information diffusion, but with a drastic reduction in terms of nodes' and network resources, i.e., generated traffic and storage space used at each node.

Acknowledgments

This work has been partially funded by the EU Commission under the H2020-INFRAIA SoBigData (654024) project. Robin I.M. Dunbar's research is funded by a European Research Council Advanced Investigator Grant.

References

[1] M. Conti, S. Das, C. Bisdikian, M. Kumar, L. M. Ni, A. Passarella, G. Roussos, G. Tröster, G. Tsudik, F. Zambonelli, Looking Ahead in Pervasive Computing: Challenges and Opportunities in the Era of Cyber-Physical Convergence, Pervasive and Mobile Computing 8 (1) (2012) 2–21. doi:http://dx.doi.org/10.1016/j.pmcj.2011.10.001.

[2] R. I. Dunbar, Grooming, Gossip, and the Evolution of Language, 1998.

[3] R. I. Dunbar, The Social Brain Hypothesis, Evolutionary Anthropology 6 (5) (1998) 178–190. doi:10.1002/(SICI)1520-6505(1998)6:5¡178::AID-EVAN5¿3.0.CO;2-8.

[4] S. G. Roberts, Constraints on Social Networks, in: Social Brain, Distributed Mind (Proceedings of the British Academy), 2010, pp. 115–134. doi:10.5871/bacad/9780197264522.001.0001.

[5] R. I. Dunbar, Constraints on the evolution of social institutions and their implications for information flow, Journal of Institutional Economics 7 (2011) 345–371. doi:10.1017/S1744137410000366.

[6] A. Sutcliffe, R. I. M. Dunbar, J. Binder, H. Arrow, Relationships and the social brain: integrating psychological and evolutionary perspectives, British Journal of Psychology 103 (2) (2012) 149–68. doi:10.1111/j.2044-8295.2011.02061.x.

[7] R. I. M. Dunbar, V. Arnaboldi, M. Conti, A. Passarella, The structure of online social networks mirrors those in the offline world, Social Networks 43 (2015) 39–47. doi:10.1016/j.socnet.2015.04.005.

[8] V. Arnaboldi, A. Guazzini, A. Passarella, Egocentric online social networks: analysis of key features and prediction of tie strength in Facebook, Computer Communications 36 (10-11) (2013) 1130–1144. doi:10.1016/j.comcom.2013.03.003.

[9] V. Arnaboldi, M. Conti, A. Passarella, F. Pezzoni, Ego Networks in Twitter: an Experimental Analysis, in: NetSciCom '13, 2013, pp. 3459–3464.

[10] P. Mac Carron, K. Kaski, R. I. Dunbar, Calling Dunbar's numbers, Social Networks 47 (2016) 151–155. arXiv:1604.02400, doi:10.1016/j.socnet.2016.06.003.

[11] R. I. Dunbar, Do online social media cut through the constraints that limit the size of offline social networks?, Royal Society Open Science 3 (1) (2016) 150292. doi:10.1098/rsos.150292.

[12] V. Arnaboldi, M. La Gala, A. Passarella, M. Conti, Information Diffusion in Distributed OSN: the Impact of Trusted Relationships, Peer-to-Peer Networking and Applications 9 (6) (2016) 1195–1208. doi:10.1007/s12083-015-0395-2.

[13] V. Arnaboldi, M. Conti, M. La Gala, A. Passarella, F. Pezzoni, Ego network structure in online social networks and its impact on information diffusion, Computer Communications 76 (2016) 26–41. doi:10.1016/j.comcom.2015.09.028.
[14] V. Arnaboldi, M. L. La Gala, A. Passarella, M. Conti, The Role of Trusted Relationships on Content Spread in Distributed Online Social Networks, in: LSDVE '14, 2014, pp. 287–298.
[15] K. Lerman, Information Is Not a Virus, and Other Consequences of Human Cognitive Limits, Future Internet 8 (2) (2016) 21. doi:10.3390/fi8020021.
[16] J. C. Cheng, L. Adamic, A. P. Dow, H. M. Kleinberg, J. Leskovec, Can Cascades Be Predicted?, in: Proceedings of the 23rd international conference on World wide web, 2014, pp. 925–936.
[17] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, S. Vigna, Four Degrees of Separation, CoRR abs/1111.4.
[18] V. Arnaboldi, A. Passarella, M. Conti, R. I. Dunbar, Online Social Networks: Human Cognitive Constraints in Facebook and Twitter Personal Graphs, Elsevier, 2015.
[19] M. S. Granovetter, The Strength of Weak Ties, The American Journal of Sociology 78 (6) (1973) 1360–1380. doi:10.2307/2776392.
[20] E. Gilbert, K. Karahalios, Predicting Tie Strength with Social Media, in: CHI '09, 2009, pp. 211–220.
[21] R. S. Burt, Structural Holes versus Network Closure as Social Capital, 2001.
[22] M. Everett, S. P. Borgatti, Ego network betweenness, Social Networks 27 (1) (2005) 31–38. doi:10.1016/j.socnet.2004.11.007.
[23] R. I. Dunbar, Coevolution of neocortex size, group size and language in humans, Behavioral and Brain Sciences 16 (1993) 681–734.
[24] W. X. Zhou, D. Sornette, R. A. Hill, R. I. M. Dunbar, Discrete hierarchical organization of social group sizes, Biological Sciences 272 (1561) (2005) 439–44. doi:10.1098/rspb.2004.2970.
[25] S. G. Roberts, R. Wilson, P. Fedurek, R. Dunbar, Individual differences and personal social network size and structure, Personality and Individual Differences 44 (4) (2008) 954–964. doi:10.1016/j.paid.2007.10.033.
[26] S. G. Roberts, R. I. Dunbar, Communication in social networks: effects of kinship, network size and emotional closeness, Personal Relationships 18 (2010) 439–452.
[27] S. G. Roberts, R. I. Dunbar, T. V. Pollet, T. Kuppens, Exploring Variation in Active Network Size: Constraints and Ego Characteristics, Social Networks 31 (2) (2009) 138–146. doi:10.1016/j.socnet.2008.12.002.
[28] R. I. M. Dunbar, M. Spoors, Social networks, support cliques and kinship, Human Nature 6 (3) (1995) 273–290.
[29] R. A. Hill, R. I. Dunbar, Social network size in humans, Human Nature 14 (1) (2003) 53–72. doi:10.1007/s12110-003-1016-y.
[30] A. G. Sutcliffe, R. I. Dunbar, D. Wang, Modelling the evolution of social structure, PLoS ONE 11 (7) (2016) 10–16. doi:10.1371/journal.pone.0158605.
[31] G. Miritello, E. Moro, R. Lara, R. Martínez-López, J. Belchamber, S. G. Roberts, R. I. Dunbar, Time as a limited resource: Communication strategy in mobile phone networks, Social Networks 35 (1) (2013) 89–95. arXiv:1301.2464, doi:10.1016/j.socnet.2013.01.003.
[32] B. Goncalves, N. Perra, A. Vespignani, Validation of Dunbar's number in Twitter conversations, arXiv.
[33] V. Arnaboldi, M. Conti, A. Passarella, R. Dunbar, Dynamics of personal social relationships in online social networks: A study on twitter, in: COSN, ACM, 2013, pp. 15–26.
[34] V. Arnaboldi, M. Conti, A. Passarella, F. Pezzoni, Analysis of Ego Network Structure in Online Social Networks, in: SocialCom '12, 2012, pp. 31–40.
[35] L. Rainie, B. Wellman, Networked: The new social operating system, The MIT Press, 2012.
[36] G. Miritello, R. Lara, M. Cebrian, E. Moro, Limited communication capacity unveils strategies for human interaction, Scientific Reports 3 (2013) 1–7. doi:10.1038/srep01950.
[37] X. Zhao, A. Sala, C. Wilson, X. Wang, S. Gaito, H. Zheng, B. Y. Zhao, Multi-scale dynamics in a massive online social network, in: IMC '12, 2012, pp. 171–184.
[38] V. Arnaboldi, M. Conti, A. Passarella, R. I. Dunbar, Dynamics of Personal Social Relationships in Online Social Networks: a Study on Twitter, in: COSN '13, 2013, pp. 15–26.
[39] B. Viswanath, A. Mislove, M. Cha, K. P. Gummadi, On the Evolution of User Interaction in Facebook, in: WOSN, 2009, pp. 37–42.
[40] A. Guille, H. Hacid, C. Favre, D. Zighed, Information diffusion in online social networks: a survey, ACM SIGMOD Record 42 (2) (2013) 17–28. doi:10.1145/2503792.2503797.
[41] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, M. Hurst, Cascading Behavior in Large Blog Graphs, arXivarXiv:0704.2803, doi:10.1.1.103.8339.
[42] F. Wang, H. Wang, K. Xu, Diffusive Logistic Model Towards Predicting Information Diffusion in Online Social Networks, in: ICDCS Workshops, 2012, pp. 133–139. arXiv:1108.0442, doi:10.1109/ICDCSW.2012.16.
[43] J. Yang, J. Leskovec, Modeling information diffusion in implicit networks, in: ICDM, 2010, pp. 599–608. arXiv:1603.02178, doi:10.1109/ICDM.2010.22.
[44] W. Galuba, K. Aberer, D. Chakraborty, Z. Despotovic, W. Kellerer, Outtweeting the Twitterers - Predicting Information Cascades in Microblogs, in: WOSN '10, 2010, pp. 3–3.
[45] K. Saito, K. Ohara, Y. Yamagishi, M. Kimura, H. Motoda, Learning diffusion probability based on node attributes in social networks, in: ISMIS, 2011, pp. 153–162.
[46] A. Guille, H. Hacid, A predictive model for the temporal dynamics of information diffusion in online social networks, in: WWW, 2012, p. 1145. doi:10.1145/2187980.2188254.
[47] E. Bakshy, I. Rosenn, C. Marlow, L. Adamic, The Role of Social Networks in Information Diffusion, in: WWW '12, 2012, pp. 519–528.
[48] J. Zhao, J. Wu, X. Feng, H. Xiong, K. Xu, Information propagation in online social networks: a tie-strength perspective, Knowledge and Information Systems 32 (3) (2011) 589–608. doi:10.1007/s10115-011-0445-x.
[49] D. Kempe, J. Kleinberg, É. Tardos, Maximizing the Spread of Influence Through a Social Network, in: KDD '03, 2003, pp. 137–146.
[50] V. Arnaboldi, M. Conti, M. La Gala, A. Passarella, F. Pezzoni, Information Diffusion in OSNs: the Impact of Nodes' Sociality, in: SAC '14, 2014, pp. 1–6.
[51] C. Wilson, A. Sala, K. P. Puttaswamy, B. Y. Zhao, Beyond Social Graphs: User Interactions in Online Social Networks and Their Implications, ACM Transactions on the Web 6 (4) (2012) 1–31. doi:10.1145/2382616.2382620.
[52] A. Passarella, R. I. Dunbar, M. Conti, F. Pezzoni, Ego network models for Future Internet social networking environments, Computer Communications 35 (18) (2012) 2201–2217. doi:10.1016/j.comcom.2012.08.003.
[53] M. Conti, A. Passarella, F. Pezzoni, A model to represent human social relationships in social network graphs, in: Social Informatics, 2012.
[54] Diaspora Project - https://diasporafoundation.org/.
[55] S. Buchegger, D. Schioberg, L. H. Vu, A. Datta, PeerSoN: P2P Social Networking - Early Experiences and Insights, in: SocialNets '09, 2009, pp. 46–52.
[56] L. A. Cutillo, R. Molva, T. Strufe, Safebook: A Privacy-Preserving Online Social Network Leveraging on Real-Life Trust, Communications Magazine, IEEE 47 (12) (2009) 94–101.
[57] M. Conti, A. D. Salve, B. Guidi, F. Pitto, L. Ricci, Trusted Dynamic Storage for Dunbar-Based P2P Online Social Networks, in: OTM '14, 2014, pp. 400–417.
[58] F. Delmastro, V. Arnaboldi, M. Conti, People-centric computing and communications in smart cities, IEEE Communications Magazine 54 (7) (2016) 122–128. doi:10.1109/MCOM.2016.7509389.
[59] M. Mordacchini, A. Passarella, M. Conti, Social Cognitive Heuristics for Adaptive Data Dissemination in Opportunistic Networks, in: WoWMoM, 2015, pp. 1–9.