A Model to Represent Human Social Relationships in Social Network Graphs

Marco Conti, Andrea Passarella, and Fabio Pezzoni

CNR-IIT, via G. Moruzzi, 1 - 56124 Pisa, Italy
{marco.conti,andrea.passarella,fabio.pezzoni}@iit.cnr.it

Abstract. *Human social relationships are a key component of emerging complex techno-social systems such as socially-centric platforms based on the interactions between humans and ICT technologies. Therefore, the models of human social relationships are fundamental to characterise these systems and study the performance of socially-centric platforms depending on the social context where they operate. The goal of this paper is presenting a generative model for building synthetic human social network graphs where the properties of social relationships are accurately reproduced. The model goes well beyond a binary approach, whereby edges between nodes, if existing, are all of the same type. It sets the properties of each social link, by incorporating fundamental results from the anthropology literature. The synthetic networks it generates accurately reproduce both the macroscopic structure (e.g., its diameter and clustering coefficient), and the microscopic structure (e.g., the properties of the tie strength of individual social links) of human social networks. We compare generated networks with a large-scale social network data set, validating that the model is able to produce graphs with the same structural properties of human-social-network graphs. Moreover, we characterise the impact of the model parameters on the synthetic graph properties.*

Keywords: social networks, human behaviour, modelling, simulations.

1 Introduction

In the last decade the proliferation of personal mobile devices, e.g. the smartphones, led to the emergence of electronic pervasive social networks which are drastically changing the way the information is circulating. In particular there is a convergence between the cyber/virtual and the physical world. Indeed, content generated in the physical space produces outcomes in the cyber/virtual world and, similarly, information generated in the cyber space has immediate influence on the physical world. At the core of this convergence there are humans which, through their devices, transfer the information between the physical and the cyber space in both directions. The analysis of the human social behaviour is therefore becoming fundamental for the development of socially centric platforms [1], whereby the properties of the social relationships between users are taken into account in the core design of the communication algorithms.

In addition, information technologies can also be used as tools to generate simulated social environments where properties of human social relationships can be studied "in vitro", under controllable parameters. For example, accurate models of human social networks can be used to study information dissemination or opinion spreading at large scale and under a range of parameters' values.

In this work we present a model for the generation of synthetic social networks whose structure reproduces the main properties of human social networks. It starts from the model presented in [2] which is able to generate single ego networks (a simple form of social network) based on well-known results in the field of anthropology. We extend the original model in order generate complete social networks formed by interconnect ego networks. With this purpose, the model relies on well-known properties in the social networks literature, such as the "triadic closure", the presence of bridges and geographical constraints [3, 4]. The parameters of the model permit to generate different social networks tuning the geographical constraints and changing the criteria the individuals use to create new social relationships. Experimental results demonstrate that generated networks accurately match the properties of human social networks. Specifically, we show that our model is able to reproduce both macroscopic properties of the network, such as its diameter and its clustering coefficient, but also microscopic properties, such as the strength of the tie of individual social links, and the correlation between the tie strength of different social links.

The use of this model for generating synthetic social network has several practical applications. On the one hand, it is a tool for accurately studying processes of social interaction via simulations. For example, it is possible to analyse variations of the information diffusion process using different settings of the model parameters. On the other hand, the model permits the development and the performance evaluation of algorithms and protocols for socially centric platforms and systems.

The remainder of this paper is organised as follows: in Section 2 we give an overview of the results regarding human social networks; in Section 3 we summarise the model for the generation of single ego networks and then we introduce the new model for the generation of complete social networks; in Section 4 we validate our model comparing different generated networks with a real human network; finally, in Section 5, we draw the main conclusions of our work.

2 Background and Related Work

The study concerning the composition and the structure of human social networks are arousing the interest of an increasing number of researchers in many different fields [2, 3, 5–12]. Significant attention has been devoted to ego networks, which are social networks between an individual (*ego*) and the other people (*alters*) the ego has a social relationship with [5]. Despite being small-size networks, ego networks are important as they permit to fully characterise the properties of social links between individuals.

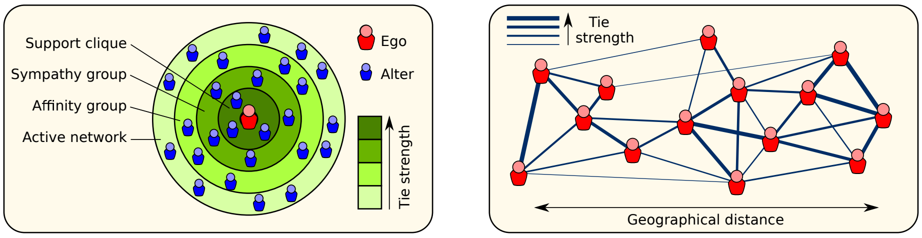

Fig. 1: (a) Ego network structure; (b) Complete social network.

One of the main results about ego networks is that their structure consists of a series of concentric layers of acquaintanceship with increasing size [6]. Based on data collected on real human networks, Dunbar et al. [12] identify four layers: "*support clique*", "*sympathy group*", "*affinity group*" and "*active network*" (the whole ego network) with average sizes of ~ 5, ~ 12, ~ 35 and ~ 150 respectively. The sizes are evaluated considering that layers are inclusive. Sometimes in this paper, we use the term *external part* of a layer in order to refer to the part of the layer not overlapped with its inner levels (e.g., for the sympathy group it is the part of the layer not overlapped with the support clique). Going from inner to outer layers, while the number of alters increases, the strength of the social tie between the ego and the alters diminishes. This means that, typically, an ego has few very strong social relationships in the support clique and a lot of weak ties in the active network (external part). The hierarchical structure of an ego network is depicted in Figure 1(a).

It has been shown that this hierarchical structure and the typical sizes of the layers are related to the level of *emotional closeness* (the strength of a social tie) and the cognitive resources humans allocate to social relationships. Intuitively, maintaining a social relationship has a cognitive cost (e.g., due to spending time together, remembering facts about the alter, etc). As the total "cognitive capacity" humans devote to social relationships is limited, the sizes of the layers are also limited [8, 9]. Other results regard the composition of each layer of the ego network with respect to the gender of the alters and to the family relationships [8, 10]. The authors of [2] define a model which allows to generate synthetic ego network graphs that satisfy all these properties. The procedure for the generation of these graphs is summarised in Section 3.1.

While being a very important model for studying certain properties of social relationships, ego networks alone cannot provide a complete representation of human social networks. Indeed they do not capture the mutual relationships between the alters or, in other words, the correlation between different ego networks. This gap can be filled connecting ego networks together in order to form a complete social network as shown in Figure 1(b). Due to the high complexity of complete social networks, the characterisation of their properties is way less advanced than that of ego networks. To the best of our knowledge, three main properties have been experimentally characterised in the literature, i.e. (i) triadic closure, (ii) the presence of bridges, and (iii) the dependence of social links on geographical distance.

Properties (i) and (ii) were investigated by Granovetter in [3]. In the paper, the author defines the *triadic closure* as a property of the social networks for which, if a strong social tie exists between two pairs of nodes A-B and B-C, there is, with a high probability, a tie between the nodes C-A which closes the triangle. The links in social networks that do not take part in triangles are called "bridges" and, according to the study in [3], they are mainly weak ties. Bridges have an important role in the social network structure as they connect socially distant parts of the network enabling to reach people and information not accessible via strong ties [3]. The presence of bridges leads the diameter of the network to be short, as in the results of the Milgram experiment [11]. At the same time, the triadic closure property guarantees a high level of clusterisation. For these reasons, human social networks can be classified as *small- world networks*, according to the definition given by Watts and Strogatz [13].

The presence of *geographical constraints* (iii) is another key factor in the formation of human social networks. Indeed, for each person, it is more likely to have a social relationship with an individual who lives close to him, than to have a tie with a person who lives far away. This hypothesis is verified experimentally by Onnella et al. in [4]. They analysed a huge data set of social interactions based on mobile phone calls in which each user is tagged with the geographical position where she probably lives. Plotting the frequencies of social ties between users which live at different distances, it emerges that the decay of the tie probability follows a power-law of the form $P(d) \sim d^{-\alpha}$, where d is the geographical distance and α is the power-law exponent. Using the maximum-likelihood method, the authors estimate $\alpha = 1.5$ [4].

In the last five years, thanks to the advent of online social networks (OSNs), the analysis of large social network graphs became more affordable. Indeed most of the recent work in social network analysis focuses on the characterisation of the global properties of a specific OSN, such as Facebook [14–16] and Twitter [17, 18]. Some important results were obtained, e.g. the validation of the "small-world property" [14], the evidence of the Dunbar's number [17] and the discovery of the power-law distribution of the degree [15]. However, these results are relevant only for the virtual environment since they are strictly related to the particular graph considered. In addition, these analyses and the resulting network models typically do not pay sufficient attention to microscopic features of social links, such as the associated tie strength, but use a binary model where links either exist or not exist (i.e., unweighted graphs).

In this work we define an original approach to social network analysis, by developing a model for the generation of human social networks which, to the best of our knowledge, reproduces the key properties of human social network highlighted in the anthropology literature. In contrast with legacy studies on OSNs we take into account the social aspects which characterise the human social networks, such as the strength of the ties, the cognitive resource consumption of the individuals and the correlation between the strength of ties between different users.

3 The Model

The model described in this section is defined by an iterative procedure able to generate synthetic social network graphs which exhibit the typical features of human social networks described in Section 2.

The procedure operates on two distinct levels of the network structure: the *local level*, in which the ego networks are generated, and the *global level*, in which the ego networks are opportunely connected to form a complete social network. Based on these distinct levels, we can consider our model as the union of two different models: a *single-ego model* and a *multi-ego model* respectively.

The single-ego model is based on the work in [2] which we summarise in Section 3.1. The multi-ego model, which relies on the concepts of triadic closures, bridges and geographical constraints, is described in detail in Section 3.2.

3.1 Single-Ego Model

The model assumes that each ego has a finite budget of cognitive resources for social relationships, expressed as the total time the ego devotes to social interactions. The algorithm adds social links to an ego network, associated with the time devoted by the ego to that particular relationship. The ego network is completed when the ego's total budget is over. The model considers a three-level structure in which layers are called "*support clique*", "*sympathy group*" and "*active network*" with average size respectively 4.6, 14.3 and 132.5 (reference values are given in [6]). This structure differs from ego network structure defined in Section 2 by the absence of the "affinity group" layer. This is justified in [2] by the lack of results about its properties currently available in literature.

The algorithm initialises each ego i with a budget of time bdg, the size of the sympathy group s_{sym} and the size of the support clique s_{sup}. Each of these values is drawn from a carefully defined density function (f_B, f_S and f_W respectively). After the initialisation, the algorithm starts creating new social ties which are characterised by a certain level of emotional closeness, extracted from a density function f_E. The level of emotional closeness is subsequently converted into time by a conversion function h, and then subtracted from the residual time budget bdg[1]. New social relationships are first included in the support clique layer until it reaches the target size, subsequently, in a similar fashion, they are included in the sympathy group (external part). For the external part of the outermost layer, the algorithm adds new social ties until the budget of time is totally exhausted.

Definitions of the density functions f_B, f_S, f_W, f_E and of the conversion function h, summarised in Table 1, are directly obtained from [2].

[1] Note that the model associates a level of emotional closeness to social ties, instead of directly associating a time budget, as the former is the typical way of characterising the strength of social ties in the anthropology literature [8, 9].

Table 1: Functions definition.

Function	Description	Definition
f_B	Time spent by egos in social activity	$Gamma(205, 8.5264)$
f_S	Sympathy group size	$Gamma(4.1, 3.49)$
f_W	Ratio between sympathy gr. and support cl. sizes	$Normal(0.3217, 0.1608)$
f_E	Emotional closeness level[a]	$Normal(0.419, 0.237)$
h	Emotional closeness → Time conversion function[b]	$h(e) = 117.18^e$

[a] We merged together the functions defined in [2] for kin and non-kin. The limits of the intervals of emotional closeness are: $\mathtt{low}_{\mathrm{sup}} = 0.8337$ and $\mathtt{low}_{\mathrm{sym}} = 0.71$.
[b] Calculated with the method described in [2] considering f_E.

3.2 Multi-Ego Model

The multi-ego model is designed in order generate complete human social networks, in which each node represents an individual whose ego network follows the model described in Section 3.1. In the multi-ego model a node is part of several ego networks with different roles. In this section we first present the high-level strategies the model follows, then we describe the algorithm in detail.

The model considers a human social network as a large group of individuals which are interconnected by social links. Intuitively, the procedure defined by the single-ego model can be applied to each of these individuals in order to generate its ego network. However, applying the single-ego procedure, we have to take into account that each new social link an individual adds to its ego network, also alters the ego network of the other individual involved in the relationship. This means checking, upon creation of a new link, that the properties of the involved ego networks are preserved. In detail, we have to check that (i) the size of the support clique, (ii) the size of the sympathy group, and (iii) the total budget of time remain consistent. Moreover, in order to generate complete ego networks we have to take into account the additional properties described in Section 2, i.e. triadic closure, presence of bridges and geographical constraints.

A new social link can be established either exploiting the triadic closure property or creating a bridge. The strategy to be used is randomly selected based on a given probability. In case the triadic closure strategy is selected, the procedure tries to close a triangle, that is, given an origin node, it selects a node at a distance of 2 hops as link's destination, favouring strong tie hops. On the contrary, in case the procedure follows the bridge creation strategy, the destination node is chosen randomly. In both cases geographical constraints have to be respected. In order to do this, we incorporate geographical information into the nodes, associating to them random locations in a virtual space. Whatever strategy to create links is selected, the model guarantees that the probability to have a social link between two nodes is proportional to a power law of the distance between them. Remember this is consistent with empirical results in the literature [4].

```
1: procedure CreateSocialNetwork(n, p, f_D, f_B, f_S, f_W, f_E, h)
2:     for i ← 1, n do
3:         i ← CreateEgo(f_B, f_S, f_W)
4:         i.pos ← ExtractFrom(Uniform(−1, 1))
5:         V ← V + i
6:     end for
7:     for all layer l ∈ {sup, sym, net} do                ▷ maintaining the ordering
8:         while Open(V, l) is not empty do
9:             i ← random select in Open(V, l)
10:            if Rand() < p then
11:                j ← ClosureSelect(i, f_D, Open (V, l))
12:            else
13:                j ← BridgeSelect(i, f_D, Open (V, l))
14:            end if
15:            r ← NewSocialLink(i, j)
16:            r.e ← ExtractFrom(f_E in (low_l, up_l))
17:            update E, i.size, j.size, i.dbg and j.bdg
18:        end while
19:    end for
20:    return V, E
21: end procedure
```

Fig. 2: Multi-ego model's algorithm.

Algorithm. The pseudo-code of the algorithm used for generating synthetic human social network graphs is shown in Figure 2. The input required by the algorithm consists of: (i) the number of nodes in the network n; (ii) the probability p to create a new social link using the triadic closure property rather than creating a bridge; (iii) the power-law distribution function f_D which gives the probability to establish a social link between nodes at a specific distance; (iv) the parameters used to define the structure of the single ego networks f_B, f_S, f_W, f_E, h, as required by the single-ego model (see Section 3.1).

In the first part of the algorithm we create and initialise each node i in the network as an ego (lines 2-6). For each node we first call the procedure CreateEgo which sets the size of the sympathy group $i.s_{\mathrm{sym}}$ and the size of the support clique $i.s_{\mathrm{sup}}$. It also assigns the budget of time $i.bdg$ and initialises the counter $i.size$ which is then used to keep track of the total size of the ego network (line 3). We also assign a geographical position of the ego ($i.pos$) which is randomly selected in a given space which, without loss of generality, we assume mono-dimensional, circular and included in the interval between -1 and 1. This definition guarantees that the distance between any pair of nodes is between 0 and 1 (line 4). Finally, each generated ego is included in the set V (line 5).

After the initialisation of the egos, we start adding social links to the network. First, we create all the social links belonging to all the support cliques, then we continue with the sympathy groups (external part), and finally we add the links of the active networks (external part) (line 7-17). Given the layer l we are populating, the creation of a new social link between two nodes i and j starts with the selection of the node i, drawn randomly from the nodes labelled as *open*

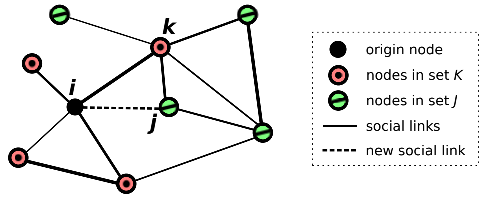

Fig. 3: Triadic closure strategy.

(line 9). An "open" node is an ego whose population of the current layer l is not yet completed[2]. The selection of the nodes involved in a new social link from the open node set $\textsc{Open}\,(V,l)$ guarantees the preservation of the ego network properties. The fundamental part of the algorithm is the selection of node j. We use two different strategies: (i) the *triadic closure* mechanism (procedure ClosureSelect) and (ii) the *bridging* (procedure BridgeSelect). The former strategy is chosen with a probability given by the parameter p, while the latter with probability $1-p$ (lines 10-14).

The *bridging*, i.e. the creation of a bridge, is the simplest strategy. We extract a node j from the open egos in the network for the current layer l, excluding the nodes already connected to i, taking into account the geographical constraints. The probability to select a node j is thus proportional to the value of the power-law function f_D (discussed in detail at the end of the section), given the distance $dist(i,j)$ between i and j. Formally,

$$P(j) \propto f_D(dist(i,j)) \qquad j \in \textsc{Open}\,(V,l) - Nei(i) - i \tag{1}$$

where $Nei(i)$ is the set of one-hop neighbours of node i.

If each node in the network, not connected to node i, is *closed* (not open), node j can not be selected. In this case node i is forced to be closed. We have experimentally checked that this circumstance occurs just in a negligible number of cases and that the overall results are not affected.

Using the *triadic closure* strategy, represented in Figure 3, we first select the set K of the neighbours of i. From this set, we extract an intermediate node k with a probability that is proportional to the tie strength e_{ik} between i and k multiplied, in order to satisfy the geographical constraints, by a function of the distance $dist(i,k)$ (Equation 2). Given the intermediate node k and the current layer l, we define the set J as the set of open neighbours of k, with respect to l, excluded node i and its neighbours. From the set J we extract node j using the same method used for the selection of node k, considering the social relationship between k and j (Equation 3).

[2] In case the current layer l is the support clique or the sympathy group, an ego i is open if its ego network size $i.size$ has not reached the thresholds $i.s_{\text{sup}}$ or $i.s_{\text{sym}}$ respectively. In case l is the active network, i is open if it has not exhausted its time budget $i.bdg$.

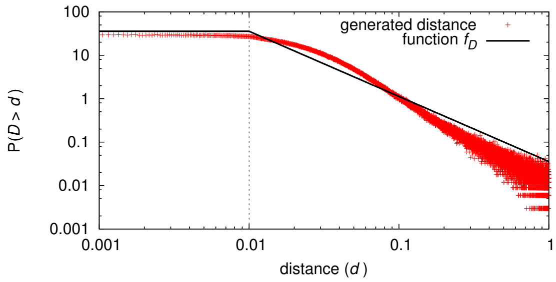

Fig. 4: PDF of the generated distance and the function f_D ($\alpha = 1.5$, $d_{min} = 0.01$)

$$P(k) \propto e_{ik} \cdot \sqrt{f_D(dist(i,k))} \qquad k \in K = Nei(i) \tag{2}$$

$$P(j) \propto e_{kj} \cdot \sqrt{f_D(dist(k,j))} \qquad j \in J = Nei(k) \cap \textsc{Open}\,(V,l) - i - K \tag{3}$$

If the set J is empty we go a step backward and we select a different node k. If, for each k chosen, it is not possible to define a non-empty set J, the procedure fails and the algorithm recovers selecting j using the bridging. Bridging is also used in case node i has not neighbours, i.e. the set K is empty.

The function of the distance we use in Equations 2 and 3 is defined as the square root of the function f_D. This definition guarantees that the geographical distance between connected nodes in the final network follows the power-law rule defined in f_D. In Figure 4 we show a comparison between a given function f_D and the geographical distances obtained using this algorithm.

After the selection of node j, a new social link r between nodes i and j is created (line 15). Its emotional closeness $r.e$ is extracted from the density function f_E in the same manner as in the single-ego model (line 16). Then, we update the network adding the new social relationship r to the set of links E. We also update the egos i and j, in terms of the ego network sizes ($i.size$ and $j.size$ respectively) and of the residual budget of time ($i.dbg$ and $j.bdg$ respectively) (line 17). It is worth noting that this update can determine the transition of a node from the open to the closed state, with respect to the current layer l.

For each layer l, we generate and add new social links until there are open nodes available. When the set of the open nodes is empty, the procedure switches to the next layer until all the three layers are completed.

Function f_D According to the results presented in [4] and summarised in section 2 the probability of contact between two users at a certain distance follows a power-law of the form $P(d) \propto d^{-\alpha}$. In order to obtain a related probability density function f_D we have to introduce a thresholds d_{min} from which the power-law hods. Moreover it has to be defined for the range of values of d, which is the interval $(0, 1)$. The function, shown in Figure 4, is thus defined as:

$$f_D(d) \propto \begin{cases} d_{min}^{-\alpha} & \text{for } 0 < d < d_{min} \\ d^{-\alpha} & \text{for } d_{min} < d < 1 \end{cases} \tag{4}$$

Experimental results in [4] suggest that $\alpha = 1.5$. On the contrary, a value for d_{min} cannot be set in general since it strongly depends on the geographical space we consider and on the geographical distribution of the sampled population. Note that, given the number n of nodes in the network, since they are equally distributed in the space, $n \cdot d_{min}$ is the average number of nodes within the distance d_{min} from any given position. Thus, given a node in the network, the closest $n \cdot d_{min}$ nodes (on average) have the same highest-probability to be selected as destination of a social link. This parameter impacts on the clustering coefficient of the network, as we highlight in section 4.2.

4 Model Validation and Properties of Generated Graphs

In this section we validate our model comparing the synthetic social networks it generates with a real social network. In Section 4.1 we describe the real social network we consider for the validation. In Section 4.2 we compare the results with the properties of the reference network and we highlight how key properties of the generated networks depend on the model parameters.

4.1 Reference Network

The reference network we use for the validation of our model is obtained from a large data set crawled from a Facebook regional network on April 2008[3]. As we discuss in [19], the analysis of this data set, opportunely processed, shows that it shares similar properties with respect to those observed in other types of human social networks, and thus it can be used as a representative network to validate our model. Note that the network resulting from this data set is of a much larger scale with respect to the ones typically analysed in the anthropology literature. It contains more than 23 million social links (Facebook friendships), involving more than 3 million users. For each social link, the data set provides the number of social interactions occurred between the users. A social interaction can be either a wall post or a photo comment. The complete analysis of this data set is available in [19]. Hereafter, we summarise the key outcomes of this analysis that are then used to validate our model.

From the original data set, some users have been dismissed since they were not considered relevant, either for having too few interactions, or because they had joined Facebook just before the beginning of the data collection period. As discussed in [19] both cases can lead to biased representations of ego networks. The new data set obtained from the selection of relevant egos and the social links between them contains $90,925$ users and $1,264,658$ social links.

As described in [19], it is possible to extract from the data set the frequency of interaction between users. Since there are evidences of a strong correlation between the interaction frequency and the strength of the social tie [8], we can

[3] This data set is publicly available for research at `http://current.cs.ucsb.edu/facebook/`, referred as "Anonymous regional network A".

Table 2: Structural properties of the reference and generated networks.

	reference network	$\boldsymbol{p} = 0.8$ $\boldsymbol{d_{min}} = \frac{250}{n}$	$\boldsymbol{p} = 0.8$ $\boldsymbol{d_{min}} = \frac{500}{n}$	$\boldsymbol{p} = 0.8$ $\boldsymbol{d_{min}} = \frac{1,000}{n}$	$\boldsymbol{p} = 0.5$ $\boldsymbol{d_{min}} = \frac{500}{n}$
mean degree	27.82	133.91	133.94	134.00	133.86
avg. shortest path	4.06	3.40	3.26	3.11	3.12
clustering coefficient	.109	.152	.108	.085	.079
Jaccard (global)	.038 [.001]	.060 [.001]	.040 [.001]	.030 [.001]	.030 [.000]
Jaccard (support cl.)	.069 [.001]	.084 [.001]	.071 [.001]	.064 [.001]	.042 [.001]
Jaccard (sympathy gr.)	.056 [.001]	.073 [.001]	.059 [.001]	.053 [.001]	.036 [.000]
Jaccard (affinity gr.)	.042 [.001]	-	-	-	-
Jaccard (active net.)	.031 [.001]	.059 [.001]	.037 [.000]	.025 [.000]	.030 [.000]

consider these frequencies define the hierarchical structure of ego networks. Authors in [19] show that 4 clusters, corresponding to the typical layers of ego networks highlighted by Dunbar [12], can be identified also in Facebook ego networks.

Relevant properties of the reference network are reported in the second column of Table 2. The high clustering coefficient (with respect to random networks) and the short average path length prove that the reference network is "small-world". Analysing the properties summarised in the table we have to take into account that, for technical reasons (e.g. the discard of not relevant nodes), the data set captures just a random sub-sample of the social links on the crawled Facebook networks and some of the indexes are influenced by the sampling, i.e. the average degree and the average path length. If we had the complete network, we would most likely find a higher average degree and a shorter path length. On the contrary, the clustering coefficient of a network preserves its value independently of the considered random sub-sample [20].

We use the Jaccard coefficient to estimate the similarity of the neighbourhoods of two adjacent nodes, that is to say the ego networks of two socially tied individuals. This is a very important index, as it describes the correlation between different ego networks. Capturing this aspect is one of the key goals of our model. The Jaccard coefficient for two sets A and B is defined as $J(A,B) = \frac{|A \cap B|}{|A \cup B|}$ and it is also not biased by random sub sampling[4]. Since computing the Jaccard coefficient between the end-points of each social link in the network requires huge computational efforts, we estimate its average value considering the pairs of end-points of a sample of $10,000$ edges randomly extracted from the network. The estimated average Jaccard coefficient (global) is reported in Table 2 (computed with 95% confidence level). According to this result, considering two socially connected individuals, their common acquaintances are, on average, 4% of the union of their acquaintances. Intuitively, individuals connected by strong ties

[4] This can be easily seen observing that random sampling proportionally affects both the union and the intersection sets.

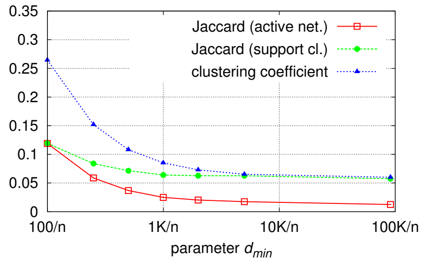

Fig. 5: Clustering coefficient and Jaccard indexes for different (a) d_{min} (with $p = .8$) and (b) p (with $d_{min} = 500/n$). 95% confidence intervals < 0.001.

should have a higher ego network similarity than individuals connected by weak ties. In order to verify this intuition, we sampled $10,000$ edges for each layer of the ego networks (external part) and computed the Jaccard coefficient between the ego networks of the nodes at the endpoints of the links. As expected, results, reported in Table 2, confirm that the similarity is higher for inner layers and lower for outer layers. Specifically, it drops from about 7% for the support clique to about 3% for the active network.

4.2 Results

The majority of the parameters for the model described in Section 3 are directly inferred from the social-anthropological literature as discussed in Section 2. The only parameters we can set in order to conduct experiments are: (i) the number of nodes in the network n; (ii) the probability of selecting the "triadic closure" strategy, and (iii) the minimum distance d_{min} for f_D. In our experiments we choose to set $n = 90,925$, which is the number of nodes in the reference network, while we use different values for the parameters p and d_{min}. The main properties of the generated network are reported in Table 2. Note that generated networks do not consider the presence of the "affinity group" layer (see Section 3.1) which we can assume to be merged with the "active network" layer.

The values of the parameters that allow us to best match the properties of the reference networks are $p = .8$ and $d_{min} = 500/n$ (fourth column of the table). These values mean that 80% of the social relationships are established through the triadic closure mechanism, rather than creating a bridge, and that, given a node, the 500 closest nodes (on average) have the same highest-probability to be selected as link's destination. Results show a strikingly similarity of the social structures between the reference network and the graph generated though the model. Indeed, both networks have the same clustering coefficient and similar Jaccard indexes for the different ego network layers. Note that discrepancies in the mean degrees and in the average shortest path length are due to the sub-sampling of the reference network. Remember that, as shown in [2], apart from these results for the global network, the use of the single-ego model (see Section 3.1) guarantees that well-known ego network properties are also satisfied. They are the size distribution of the network and of the single layers, the correlation between the layer dimensions and the distribution of the emotional closeness level.

In Table 2 we report the properties of the networks obtained with $d_{min} = 250/n$ (third column of the table) and $d_{min} = 1,000/n$ (fifth column of the table), maintaining $p = .8$. Moreover, Figure 5 (a) shows the clustering coefficient and the Jaccard index computed between pairs of strongly-tied egos (i.e. belonging to each other support clique) and weekly-tied egos (belonging to each other active network). Results show that reducing d_{min} the clustering coefficient and the similarity indexes increase for all layers of the network. Intuitively, this is because with smaller d_{min} the set of nodes selected with highest probability by an ego (those at a maximum distance of d_{min}) is smaller, and geographically very close to the ego. This leads to higher clustering (and similarity).

Similarly to the geographical constraints, also the variation of the parameter p influences the structure of the network. As shown in the last column of the table and in Figure 5 (b), if we diminish the value of p, the clustering coefficient and the similarity indexes decrease. This is expected as the number of links established though the bridging increases, and the bridging mechanism alone leads to the generation of random networks without clusters of socially connected nodes. Note in particular that when $p = 0$ (corresponding to a network without triadic closures) the Jaccard indices in Figure 5 (b) are the same, as in a network without triadic closures the correlation between social links do not depend on the strength of the links anymore.

5 Conclusions

In this work we define a new model for the generation of social network graphs, significantly extending the ego network model presented in [2]. We introduce different strategies to combine ego networks in order to form complete social network graphs, based on well-known properties in the field of social networks analysis i.e. (i) the "triadic closure“, (ii) the presence of bridges and (iii) the geographical constraints.

In order to validate our model, we tune the model parameters obtaining a graph with the same structural properties of a real large scale human network obtained from Facebook. Then, we analyse the effect of key parameters on the properties of the generated graphs, highlighting the impact of both geographical constraints and social constraints.

The results presented in the paper confirm that our model leads to the generation of network graphs socially consistent. This model can thus be used for analysing through large scale simulation key properties of human social networks and for the development and the validation of protocols for socially-centric platforms.

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

References

1. Conti, M., et al.: Looking ahead in pervasive computing: Challenges and opportunities in the era of cyber-physical convergence. Pervasive and Mobile Computing **8**(1) (2012) 2 – 21
2. Conti, M., Passarella, A., Pezzoni, F.: A model for the generation of social network graphs. In: World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium on a. (June 2011) 1 –6
3. Granovetter, M.: The strength of weak ties. American journal of sociology (1973) 1360–1380
4. Onnela, J.P., Arbesman, S., Gonzlez, M.C., Barabsi, A.L., Christakis, N.A.: Geographic constraints on social network groups. PLoS ONE **6**(4) (04 2011) e16939
5. Roberts, S.G., et al.: Exploring variation in active network size: Constraints and ego characteristics. Social Networks **31**(2) (2009) 138 – 146
6. Zhou, W.X., Sornette, D., Hill, R.A., Dunbar, R.I.M.: Discrete hierarchical organization of social group sizes. Proceedings of the Royal Society B: Biological Sciences **272**(1561) (2005) 439–444
7. Dunbar, R.I.M.: The social brain hypothesis. Evolutionary Anthropology: Issues, News, and Reviews **6**(5) (1998) 178–190
8. Hill, R.A., Dunbar, R.I.M.: Social network size in humans. Human Nature **14**(1) (2003) 53–72
9. Roberts, S., Dunbar, R.: Communication in social networks: Effects of kinship, network size, and emotional closeness. Personal Relationships **18**(3) (2011) 439–452
10. Dunbar, R.I.M., Spoors, M.: Social networks, support cliques, and kinship. Human Nature **6**(3) (September 1995) 273–290
11. Travers, J., Milgram, S.: An Experimental Study of the Small World Problem. Sociometry **32**(4) (December 1969) 425
12. Sutcliffe, A., Dunbar, R., Binder, J., Arrow, H.: Relationships and the social brain: Integrating psychological and evolutionary perspectives. British Journal of Psychology **103**(2) (2012) 149–168
13. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**(6684) (June 1998) 440–442
14. Ugander, J., Karrer, B., Backstrom, L., Marlow, C.: The Anatomy of the Facebook Social Graph. CoRR **abs/1111.4** (2011)
15. Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Measurement and analysis of online social networks. Proceedings of the 7th ACM SIGCOMM conference on Internet measurement IMC 07 **40**(6) (2007) 29
16. Wilson, et al.: User interactions in social networks and their implications. In: Proceedings of the 4th ACM European conference on Computer systems. EuroSys '09, New York, NY, USA, ACM (2009) 205–218
17. Gonalves, B., Perra, N., Vespignani, A.: Validation of dunbar's number in twitter conversations. CoRR **abs/1105.5170** (2011)
18. Teutle, A.: Twitter: Network properties analysis. In: Electronics, Communications and Computer (CONIELECOMP), 2010 20th International Conference on. (February 2010) 180 –186
19. Arnaboldi, V., Conti, M., Passarella, A., Pezzoni, F.: Analysis of ego network structure in online social networks. Technical Report IIT-CNR TR-10-2012 (2012) Available as `http://www.iit.cnr.it/sites/default/files/TR-10-2012.pdf`.
20. Lee, S.H., Kim, P.J., Jeong, H.: Statistical properties of sampled networks. (2009)