The role of strong and weak ties in Facebook: a community structure perspective

Emilio Ferrara

Department of Mathematics

University of Messina

V.le F. Stagno D'Alcontres 31, 98166 ME (Italy)

`eferrara@unime.it`

Pasquale De Meo, Giacomo Fiumara, Alessandro Provetti

Department of Physics, Informatics Section

University of Messina

V.le F. Stagno D'Alcontres 31, 98166 ME (Italy)

`{pdemeo,gfiumara,ale}@unime.it`

Abstract

In this paper we report our findings on the analysis of two large datasets representing the friendship structure of the well-known Facebook network. In particular, we discuss the quantitative assessment of the *strength of weak ties* Granovetter's theory, considering the problem from the perspective of the community structure of the network. We describe our findings providing some clues of the validity of this theory also for a large-scale online social network such as Facebook.

Complex networks, Social network analysis, Community structure, Strength of weak ties

1 Introduction

This work presents the results of our analysis carried out on two different large datasets representing the friendship structure of the well-known Facebook online social network. In particular, we here provide with some clues in the direction of verifying the renowned theory about (off-line) social networks and to propose a new perspective which, in our opinion, best captures the original intuition underlying so-called *weak ties* and their role in complex social networks.

The analysis and understanding of online social networks (OSNs) such as Facebook finds a theoretical and conceptual foundation in *social network analysis*, a field of *computational social sciences*. Moreover, several *computer science* challenges hold, given the size, distribution and organization (privacy, visibility rules etc.) of the data available

even to the regular user of such services. At a certain level of abstraction, online social networks might be seen as complex networks that describe interactions that are non-deterministic in nature. In such context, the analysis of large subsets of an OSN shall lead to statistically-robust measurements that are the basis of any understanding of the OSN structure and evolution. In particular, management and marketing decisions are based on such aggregate measures.

In this paper, we have been concerned with the experimental assessment of the importance, foreseen by the early work of Mark Granovetter [14] of *weak ties:* human relationships (acquaintance, loose friendship etc.) that are less binding than family and close friendship but might, according to Granovetter, yield better access to information and opportunities. Facebook is organized around the recording of just one type of relationship: friendship. Of course, Facebook friendship captures several degrees and nuances of the human relationships that are hard to separate and characterize within data analysis. However, weak ties have a clear and valuable interpretation: friendship between individuals who otherwise belong to distant areas of the friendship graph. Or, in other words, happen to have most of their other relationships in different national/linguistic/age/common experience groups. Such weak ties have strength precisely because they connect distant areas of the network, thus yielding several interesting properties, which will be discussed in the next sections.

2   Methodology

The classical definition of *strength of a social tie* has been provided by Granovetter [14]:

> *The strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie.*

This definition introduces some important features of a social tie that will be discussed later, in particular: *(i)* intensity of the connection, and *(ii)* mutuality of the relationship.

Granovetter's paper gives a formal definition of *strong* and *weak ties* by introducing the concept of *bridge*:

> *A bridge is a line in a network which provides the only path between two points. Since, in general, each person has a great many contacts, a bridge between A and B provides the only route along which information or influence can flow from any contact of A to any contact of B.*

From this definition it emerges that – at least in the context of social networks – *no strong tie is a bridge.* However, that is not sufficient to affirm that each weak tie is a bridge, but what is important is that *all bridges are weak ties.*

Granovetter's definition of *bridge* is restrictive and unsuitable for the analysis of large-scale social networks. In fact, because of some well-known features such as the *small world* effect (*i.e.*, the presence of short-paths connecting any pair of nodes) and the *scale-free degree distribution* (*i.e.*, the presence of hubs that maintain efficiently connected the network), it is unlikely to find an edge whose deletion would lead to the inability for two nodes to connect each other by means of alternative paths.

On the other hand, without loss of generality on a large scale, we can define a *shortcut bridge* as the link that connects any pair of nodes whose deletion would cause an increase of the distance between them, for example defining the distance of two nodes as the length of the shortest path linking them. Unfortunately, also this definition leads to two relevant problems. The former is due to the introduction of the concept of shortest paths; the latter is due to the possible arbitrariness given by the concept of distance between nodes. In detail, regarding the shortest paths, the computation of all pairs shortest paths has a high computational cost which makes it unfeasible even on networks of modest size – even worse if considering large social networks. Regarding the second aspect, in the context of shortest paths the distance could be considered as the number of hops required to connect two given nodes. Alternatively, it could be possible to assign a value of strength (*i.e.*, a weight) to each edge of the network and to define the shortest distance between two nodes as the cost of the cheapest path joining them[1]. In such a case, however, we do not know whether this definition of distance is better than in the previous one – or if it yields to better results – but its computation remains excessively expensive in real-world networks.

In the light of the considerations above, we suspect that the problem of discriminating weak and strong ties in a social network is not trivial, at least on a large scale. To this purpose, in the following we give a definition of weak ties from a different perspective, trying not to distort the Granovetter's original intuition.

In particular, recalling that *weak ties* are considered as loose connections between any given individual and her/his acquaintances with whom she/he seldom interacts and who belong to different areas of the social graph, we give the definition of *weak ties* as those ties that connects any pair of nodes *belonging to different communities.*

To this purpose, note that our definition is more relaxed than that provided by Granovetter. In detail, the fact that two nodes connected by a tie belong to different communities does not necessarily imply that the connection among them is a bridge, nor a shortcut bridge, since its deletion could not increase the length of the path connecting them (there could yet exist one or more paths of the same length). On the other hand, in our opinion, it is a reasonable assumption at least in the context of large social networks, since it has been proved that the edges connecting different communities are bottlenecks [21] and their iterative deletion causes the fragmentation of the network in disconnected components. One of the most important characteristics of *weak ties* is that those which

---

[1]In this context, measuring the strength of the edges in online social networks has been recently advanced by [25, 11, 27].

are bridges create more, and shorter, paths. The effect in the deletion of a weak ties would be more *disruptive* than the removal of a strong tie[2], from a community structure perspective.

2.1 Experimental Set Up

In order to assess the *strength of weak ties* theory on a large scale, first of all we carefully analyzed the features of existing online social networks, considering some requirements that come directly from Granovetter's seminal work [14]:

> *Ties discussed in this paper are assumed to be positive and symmetric. Discussion of operational measures of and weights attaching to each of the four elements is postponed to future empirical studies.*

Granovetter introduces two concepts that are crucial to understanding weak ties. The first is related to the *symmetry of the relationship* among two individuals of the network. This concept is extremely interconnected with the definition of *mutual friendship relation* which characterizes several online social networks. In detail, a friendship connection can be symmetric (i.e., mutual) if there is no directionality in the relation between two individuals – otherwise the relation is asymmetric – of which Facebook *friendship* is perhaps the best-known example.

While in real-world social networks the classification of a relation between individuals can be not trivial, online social network platforms permit to clearly and uniquely define different types of connections among users. For example, in Twitter the concept of relation between two individuals intrinsically implies a directionality. In fact, each user can be a *follower* of others, can *retweet* their *tweets* and can *mention* them. Recently, research has started on assessing the *strength of weak ties* in the context of a directed network [24, 13].

Another aspect whose consideration is also important, however, is the *weight* assigned to connections (regardless of they are directed or not). The possibility of weighting connections among users of social networks has been recently envisaged by us [6] – in particular, considering the tendency of a given connection to foster the information propagation – as well as by other authors [25, 11, 27]. Nevertheless, we deem a network that can be represented by an unweighted graph the most appropriate setting for a quantitative validation of the theory – in order not to introduce an additional parameter possibly causing bias in our evaluation.

Facebook arguably represents an ideal setting for the validation of the *strength of weak ties* theory. In fact, both of Granovetter's requirements are satisfied in the Facebook friendship network because:

---

[2]For this reason *weak ties* have been recently proved to be very effective in the diffusion of information and in the rumor spreading through social networks [3, 28].

- it is naturally represented as an *undirected graph – friendship* in Facebook is symmetric –, and

- can be represented by using an *unweighted graph*, evaluating all connections in a democratic way[3].

To sum it up, our definition of the *Facebook social graph* is simply an unweighted, undirected graph $G = \langle V, E \rangle$ where vertices $v \in V$ represent Facebook users and edges $e \in E$ represent the friendship connections among them.

In this context, we define as *weak ties* those edges that, after dividing the network structure in communities (obtaining the so-called *community structure*), connect nodes belonging to different communities. *Vice versa*, we classify as *strong ties* the intra-community edges.

2.2 Community Detection

In the formulation of our problem clearly emerges the importance of the aspect of detecting communities in the network and dividing it so that each node is assigned – at least – to one community.

The problem of clustering networks is challenging and several solutions have been suggested in literature. Particular research efforts have been recently spent in the direction of unveiling the community structure of complex networks. Due to space limitations, the material presented in this section is not exhaustive and we refer the reader to some comprehensive surveys [9].

Given a network represented by a graph $G = \langle V, E \rangle$, the *community structure* is a partition $P = \{C_1, C_2, \ldots, C_r\}$ of vertices of G such that, for each $C_i \in P$, the number of edges linking vertices in C_i is much higher than the number of edges linking a vertex of C_i with a vertex residing outside C_i. Each set C_i is called *community*.

There exist different popular paradigms to discover communities. In the following we briefly discuss the so-called *network modularity maximization strategies*.

2.2.1 Network Modularity Maximization

The *network modularity* – usually denoted as Q – is a function that evaluates the quality of a partitioning of a graph $G = \langle V, E \rangle$ [9]. The higher Q, the better the partitioning. Strategies based on the maximization of the network modularity rely on the idea that *random graphs* are not expected to exhibit a *community structure*. Therefore, given a graph G and a subgraph $C \subseteq G$, a *null model* G' associated with G is defined as a

---

[3]Of course, this is not necessarily the only valid representation of the Facebook network since it should be possible to adopt a weighted network where edge weights represent, for example, the frequency of interaction between each pair of friends.

graph having the same number of vertices and edges of G, but these edges could be distributed according to some probability distribution: for instance, in case of uniform probability we obtain the so-called *Bernoulli random graph* which yields to a Poissonian degree distribution [9].

Owing to the presence of a null model, it is easy to decide whether a subgraph $C \subseteq G$ is a community or not. In fact, since G and G' have the same set of vertices, we can consider the subgraph $C' \subseteq G'$ obtained by isolating, in G', the vertices forming C in G. As claimed before, the null model is expected not to present a community structure and, therefore, we expect that C' is *not* a community. Therefore, if the density of internal edges of C is much higher than that of C', we can conclude that C is a community.

According to these observations, the *modularity function* is defined as

$$Q = \frac{1}{2m} \sum_{i,j} (A_{ij} - P_{ij}) \, \delta(C_i, C_j)$$

where m is the total number of edges in G, A_{ij} is the adjacency matrix of G, P_{ij} is the expected number of edges between i and j in the null model[4] and $\delta(\cdot, \cdot)$ is the Kronecker symbol (*i.e.*, $\delta(C_i, C_j) = 1$ if and only if $C_i = C_j$ and 0 otherwise).

Various null models are, in principle, allowed and, for each of them, we could derive a suitable expression for P_{ij}. The most common choice, however, is to assume that P_{ij} is proportional to the product of the degrees k_i and k_j of i and j respectively. According to this choice, Q can be rewritten as follows

$$Q = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i \cdot k_j}{2m} \right) \delta(C_i, C_j) \tag{1}$$

The problem of maximizing Equation 1 has been proved to be *NP-hard* and several heuristic strategies have been proposed as to date. Among them, the efficient technique called *Louvain method* (LM) [1, 5] has been adopted during our experiments and it is briefly described in the following.

2.2.2 The *Louvain method*

The *Louvain method* (LM) has been proposed in 2008 by Blondel et al. [1] and it is perhaps one of the most popular algorithms in the field of community detection. This popularity derives by the fact that LM provides excellent results even if the networks to process are very large.

The input of the algorithm is a weighted network $G = \langle V, E, W \rangle$ being W the weights associated with each edge[5]. The modularity is defined as in Equation 1, in which A_{ij} is

---

[4]Observe that P_{ij} is a real number in $[0, 1]$.
[5]Of course, in case of unweighted graphs, W is the adjacency matrix of G.

the weight of the edge linking i and j and k_i (resp., k_j) is the sum of the edges incident onto i (resp., j). Initially, each vertex i will form a community and therefore, there are as many communities as vertices in V.

LM consists of two steps which are iteratively repeated. In the first step, for each vertex i, LM considers the neighbors of i; for each neighboring vertex j, LM computes the *gain of modularity* that would take place by removing i from its community and placing it in the community of j. The vertex i is placed in the community for which this gain achieves its maximum value. Of course, if it is not possible to achieve a positive gain, the vertex i will remain in its original community. This process is applied repeatedly and sequentially for all vertices until no further improvement can be achieved. This ends the first phase.

The second step of LM generates a new weighted network G' whose vertices coincide with the communities identified during the first step. The weight of the edge linking two vertices i' and j' in G' is equal to the sum of the weights of the edges between the vertices in the communities of G corresponding to i' and j'. Once the second step has been performed, the algorithm re-applies the first step. The two steps are repeated until there are no changes in the obtained community structure.

LM has been chosen not only for its computational efficiency but also because it has got three nice properties: *(i)* it generates a hierarchy of communities and the k-th level of the hierarchy corresponds to the set of communities found after k iterations; *(ii)* even though the most time expensive part of the algorithm is the evaluation of the gain attained by moving a vertex from a community to another one, the authors provided an efficient formula to quickly compute such a gain; *(iii)* its output is stable.

2.3 Dataset

One important step in the analysis of OSNs is acquiring relevant information from the online platforms. This stage is time consuming, since it requires techniques such as Web mining and a background in statistical sampling methods. In order to investigate the *strength of weak ties* theory in the context of OSNs we adopted two large datasets already presented by Gjoka *et al.* [12], which represent two samples taken from the largest – as to the date – existing online social network: Facebook. More in detail, the datasets we adopt represent snapshots of the structure of the friendship network, also called *social graph*, among users subscribed to Facebook at the time of the sampling (April 2009). The structure of the social graph is represented by means of an undirected/unweighted graph $G = \langle V, E \rangle$ in which the set of vertices V represents social network users and the set of edges E represents friendship connections among them.

These samples have been collected by adopting two different sampling techniques, which minimize the bias introduced by the partial visit of the overall Facebook graph, whose size has been only recently estimated in 721 millions of nodes and 69 billions of edges [26].

In detail, in this study we consider two different samples collected and made publicly available in anonymized format by [12]: *(i)* Uniform sample; *(ii)* Metropolis-Hastings Random Walk sample.

The former dataset is unbiased for construction, at least in its formulation for the sampling problem for Facebook. It is obtained by using a rejection-based sampling technique, which generates an arbitrarily large list of randomly chosen user identifiers. The latter sampling method is based on Metropolis-Hastings Random Walks (MHRW). The MHRW algorithm (a general Markov Chain Montecarlo method) has been proved to work well in the context of sampling online social networks [12]. Starting from an arbitrary number of seeds (in the case of [12] the authors selected 28 Facebook users profiles), the sampling algorithm performs a MHRW moving towards vertices with lower degrees with a heightened probability with respect to following paths towards vertices with high degree. This is done on the purpose of avoiding the bias towards high degree vertices which has been proved to be introduced by sampling methods such as the breadth-first-search [16]. Regardless of the sampling method adopted to select the Facebook user to query at each step of the sampling process, the following details are retrieved: *(i)* the friendship connections among the selected user and all her/his friends on the platform; *(ii)* the geographical location of the selected user, represented by means of the regional network identifier assigned by Facebook to the last specific geographical position from which the user logged into the platform[6].

Required data have been retrieved querying the front-end of the Facebook social platform. They have been stored and then, after a cleansing phase during which the authors verified the integrity and coherency of the results, data have been released under an anonymized format, in order to preserve the privacy of the users. In the following we briefly describe the features of the social graph for the datasets provided by [12].

2.3.1 Description of the original dataset

The *Uniform sample* (hereafter, UNI) represents a social graph $G = \langle V, E \rangle$ containing 984 thousands of vertices and 72.2 millions of edges. The *MHRW sample* (henceforth, MHRW) is a social graph $G = \langle V, E \rangle$ constituted by 957 thousands of vertices and 58.4 millions of edges. In UNI, the average degree of each vertex (*i.e.*, the number of friendship connections of the given user with other users) is 95.2 while in the MHRW is 94.1. This is consistent with the overall statistics officially released by Facebook at the time of the sampling. All further details regarding UNI and MHRW are provided in [12].

---

[6]Please note that, even if in this work we do not exploit geographical information, as discussed in the conclusive section, it is the aim of ongoing research to focus on both social and geographical data of social network users.

2.3.2 Building our ad-hoc Facebook dataset

Since our purpose is to obtain a graph which reflects the actual community structure of a social network, we had to shrink the amount of friendship connection included in the datasets provided by [12]. This happens because these samples include not only those users which have been both discovered and visited during the sampling process, but also all their friends, *i.e.*, those users that represent the *frontier* of the sampling process – that have been *discovered* but not visited.

In order to build our *ad hoc* social dataset, first of all we merged the two graphs provided by [12], *i.e.*, UNI and MHRW. The amount of overlapping users between the two samples was only 4.1 thousands, thus the fused graph we obtained was constituted by about 1.9 million users. Unfortunately, a significant part of these users were connected only with friends belonging to the set of *discovered users*. For such a reason, from this graph we retained only those users for which there was existing at least one edge connecting them to another user belonging to the same set. This has been done in order to obtain a graph in which any user belonged to the list of users visited during the process of sampling (and not only discovered). The final social graph contains 613 thousands of users and 2.04 millions friendship connections among them.

The social graph built as discussed above has been exploited during the experiments as follows.

3   Experiments

Recently, several works focused on the Facebook social graph [12, 2, 26] and on its community structure [18, 8, 7], but none of them has been carried out to assess the validity of the *strength of weak ties* theory. In this section: *(i)* firstly, we characterize the node degree distribution and the size of the communities present in the Facebook *social graph*; *(ii)* secondly, we investigate presence and behavior of strong and weak ties in such a network; finally, *(iii)* we try to describe the density of weak ties among communities and the way in which the are distributed as a function of the size of the communities themselves.

3.1   Node degree and community size distribution in Facebook

Our first analysis aims at describing the distribution of node degree in Facebook. To this purpose, we adopted the *complementary cumulative distribution function* (CCDF), defined as $\hat{F}(x) = \Pr(X > x)$ – *i.e.*, the probability that a random variable X assumes values above a given x. The results are depicted in Figure 1, in which the CCDF of the probability of finding a node of given degree in the network is presented. A debate is currently ongoing in the research community to assess if this kind of behavior recalls or

not a power law in Facebook [2, 12, 26]. Regardless, we assess that the obtained degree distribution is very similar to that of the original social graph [12]. In fact, a large amount of nodes in the network have a relatively small degree, and the distribution falls off allowing the presence of a few nodes having a large degree. As a further consideration, we highlight that the average degree is ≈ 22.74, smaller than that of the original graphs [12], due to the operation of shrinking on the graph discussed in Section 2.3.2, which caused the removal of those edges linking to nodes in the frontier of the graph – and of the nodes themselves.

Then, we focused our attention on the study of the distribution of the size of discovered communities. Results are reported in Figure 2, in which we adopted once again the CCDF to describe the probability of finding a community of a certain size in the community structure of the network, unveiled by means of the *Louvain method*. The resulting distribution is well represented by a power law[7], being the *log-log* plot an almost straight line. This means that the community detection process discovered a large amount of small communities and a quickly decreasing amount of larger communities. The total number of discovered communities is $196,665$ – the largest contains $1,471$ members, with an average size of ≈ 9.

We finally recall that the presence of a power law distribution with a clear amount of small communities is important also for the evaluation of the so-called resolution limit [10]. This problem affects *modularity maximization* algorithms, such as the *Louvain method* and, depending on the topology of the network, causes the inability of the process of community detection to find communities whose size is smaller than $\sqrt{E/2}$ (*i.e.*, in our case $\approx 1,000$). We hence assessed that the community structure unveiled by the algorithm for our graph is unlikely to be affected by the resolution limit, being the most of the communities revealed smaller than that size and well distributed according to a power law.

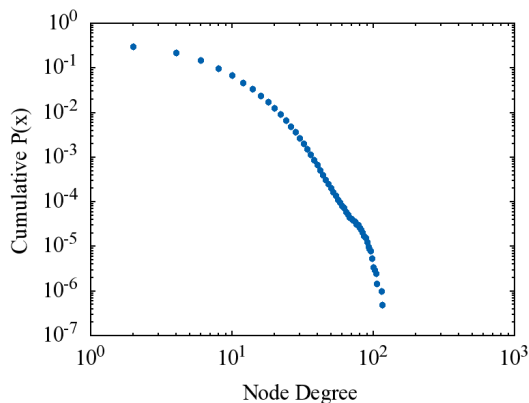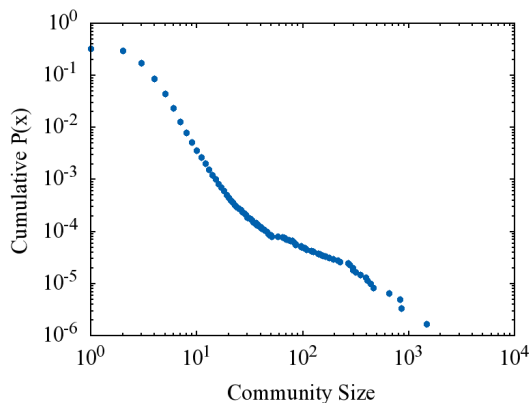Figure 1: CCDF of Facebook node degree.   Figure 2: CCDF of the size of the communities in Facebook.

---

[7]The power law distribution of the size of the communities in Facebook has been discussed in our related work [8].

3.2 Distribution and CCDF of strong and weak ties

The second experiment is devoted to understand the presence and the distribution of strong and weak ties among communities. To this purpose, we consider the community structure discussed above, classifying those edges connecting nodes belonging to different communities as weak ties, and strong ties the *vice versa*.

Intuitively, given the power law distribution of the size of communities (and, coincidentally, the power law distribution of node degrees), the number of weak ties will be much greater than the number of strong ties. Even though this effect could appear as counter-intuitive (for example, we could suppose that weak ties are much more rare than strong ties on a large scale), we should recall that some sociological theories[8] assume that individuals tend to aggregate in small communities[9], *i.e.*, the most of connections among individuals are weak ties in the Granovetter's sense – small amount of contacts, low frequency of interactions, etc.

This intuitions are reflected by analyzing Figure 3. For each node $v \in V$ of the graph $G = \langle V, E \rangle$, Figure 3 depicts the amount of strong and weak ties incident on v. It is evident that the weak ties are much more that the strong ties. The two distributions tend to behave quite similarly, but they maintain a certain constant offset which represents the ratio between strong and weak ties in this network. This ratio has been assessed in indicatively 80%-20% and carries also an important social interpretation. In fact, it is closely related to the concept of *rich club* – deriving from the renown *Pareto principle* [20] – whose validity has been recently proved for complex networks [4] (for example for Internet [29] and scientific collaboration networks [23]).

In addition, since both the distributions recall a straight line (which, in a *log-log* plot, induces to scale-free behaviors), we can assume that also the distribution of weak and strong ties could be well described by means of power laws, such as in the case of node degree and size of communities.

Considering the same problem from a different perspective, Figure 4 represents the CCDF of the probability of finding a given number of strong and weak ties in the network. From its analysis, it emerges an important difference between the behavior of the weak and the strong ties. In detail, the cumulative probability of finding a node with an increasing number of strong ties quickly decreases. Tentatively, it is possible to identify in $k \approx 5$ the tipping point from which the presence of weak ties quickly overcomes that of strong ties, making the latters less numerous in nodes with degree higher than k.

---

[8]For example, cognitive balance [19, 15], triadic closure [14] and homophily [17].

[9]According to these theories, we can explain that the intensity of human relations is very tight in small groups of individuals, and decreases towards individuals belonging to distant communities.

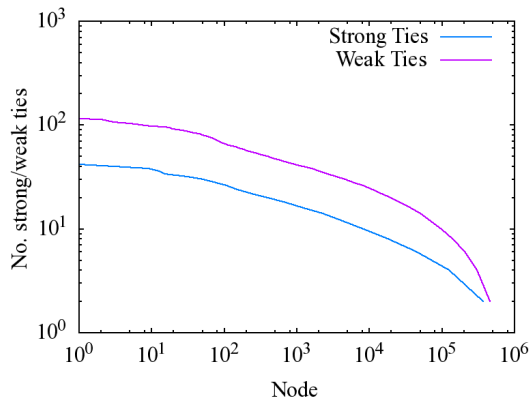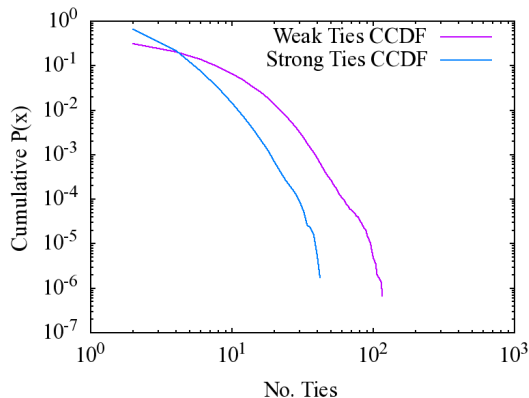Figure 3: Distribution of strong vs. weak ties in Facebook.   Figure 4: CCDF of strong vs. weak ties in Facebook.

3.3 Density of weak ties among communities and link fraction

The last experiment discussed in this paper is devoted to understanding the density of weak ties connecting communities in Facebook. In particular, we are interested in defining to what extent a weak tie links community of comparable or different size. To do so, we considered each weak tie in the network and we computed the size of the community to which the *source node* of the weak tie belongs to. Similarly, we computed the size of the *target* community[10].

Figure 5 represents a density map of the distribution of weak ties among communities. First, we highlight that the map is symmetric with respect to the diagonal, according to the fact that the graph is undirected and each weak tie is counted twice, once for each end-vertex. From the analysis of this figure, it clearly emerges that the weak ties mainly connects nodes belonging to small communities. To a certain extant, this could be intuitive since the number of communities of small size, according to their power law distribution, is much greater than the number of large communities. On the other hand, it is an important assessment since similar results have been recently described for Twitter [13]. Thus, it emerges that one of the roles of weak ties is to connect small communities of acquaintances which are not that close to belong to the same community but, on the other hand, are somehow proficiently in contact.

As for further analysis, we carried out another investigation oriented to the evaluation of the amount of weak ties that fall in each given community with respect to its size. The results of this assessment are reported in Figure 6. The interpretation of this plot is the following: on the y-axis it is represented the fraction of weak ties per community as a function of the size of the community itself, reported on the x-axis. It emerges that also the distribution of the link fraction against the size of the communities resembles a power law.

---

[10]We recall that, being the network model adopted undirected, the meaning of source and target node is only instrumental to identify the end-vertex of each given edge.

Indeed, this result is different from that recently proved for Twitter [13], in which a Gaussian-like distribution has been discovered. This is probably due to the intrinsic characteristics of the networks, that are topologically dissimilar (*i.e.*, Twitter is represented by a directed graph with multiple type of edges) and also the interpretation itself of social tie is different. In fact, Twitter represents in a way *hierarchical connections* – in the form of *follower* and *followed* users – while Facebook tries to reflects a friendship social structure which better represents the community structure of real social networks.

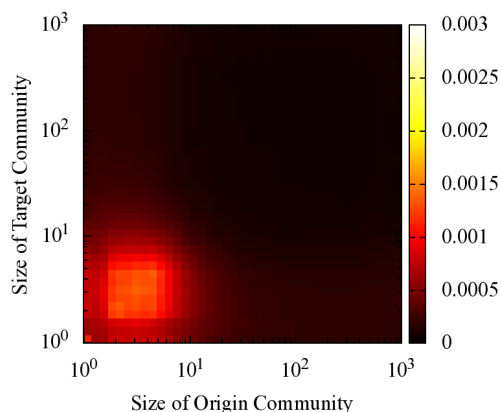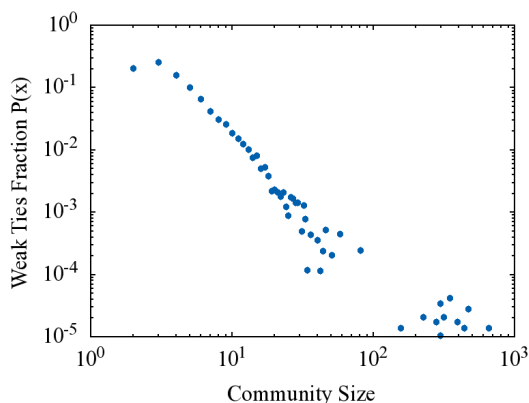Figure 5: Density of weak ties among communities.

Figure 6: Link fraction as a function of the community size.

4   Conclusions

In this paper we presented a quantitative analysis, carried out on a large sample of the Facebook online social network, to assess the validity of the *strength of weak ties* Granovetter's theory [14]. According to the formulation presented in his seminal paper, we analyzed the presence and the role of strong and weak ties in Facebook with respect to the community structure of the network. Our experimentation provided with some clues of role and importance of weak ties. We characterized their overall statistical distribution, as a function of the size of the communities and the density of weak ties among communities.

As for future works, we present two relevant ongoing research efforts related to this research. The first is the investigation of the applicability of a network weighting strategy so that the strength of ties can be computed according to a given rationale, for example the ability of each link to spread information. In fact, as previously remarked, an important aspect of *weak ties* is their ability in enhancing information diffusion through the social network. According to this idea, we intend to adopt a novel method of weighting edges well suited for social networks [6] to identify and study strong and weak ties.

Another ongoing research is related to exploiting the geographical data we already col-

lected, regarding the physical location of users of Facebook. In fact, to study the effect of strong and weak ties in the society, is it known that a relevant additional source of information is represented by the geographical distribution of individuals [22]. We aim at merging information from different graphs (*e.g.*, *social* and *geographical*) and exploiting them to get additional insights about the role of physical and virtual distances. For example, we suppose that strong ties could reflect relations characterized by physical closeness, while weak ties could be more appropriate to represent connections among physically distant individuals.

References

[1] V. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008:P10008, 2008.

[2] S. Catanese, P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti. Crawling facebook for social network analysis purposes. In *Proc. International Conference On Web Intelligence, Mining And Semantics*, pages 52:1–52:8, 25-27 May 2011.

[3] D. Centola. The spread of behavior in an online social network experiment. *Science*, 329(5996):1194, 2010.

[4] V. Colizza, A. Flammini, M. Serrano, and A. Vespignani. Detecting rich-club ordering in complex networks. *Nature Physics*, 2(2):110–115, 2006.

[5] P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti. Generalized louvain method for community detection in large networks. In *Proc. 11th International Conference On Intelligent Systems Design And Applications*, pages 88–93. IEEE, 2011.

[6] P. De Meo, E. Ferrara, G. Fiumara, and A. Ricciardello. A novel measure of edge centrality in social networks. *Knowledge-based Systems. doi: 10.1016/j.knosys.2012.01.007 (2012)*.

[7] E. Ferrara. A large-scale community structure analysis in facebook. In *arXiv:1106.2503v3*, 2011.

[8] E. Ferrara. Community structure discovery in facebook. *International Journal of Social Network Mining*, 1(1):67–90, 2012.

[9] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.

[10] S. Fortunato and M. Barthelemy. Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36, 2007.

[11] E. Gilbert and K. Karahalios. Predicting tie strength with social media. In *Proc. 27th international conference on Human factors in computing systems*, pages 211–220. ACM, 2009.

[12] M. Gjoka, M. Kurant, C. Butts, and A. Markopoulou. Practical recommendations on crawling online social networks. *Selected Areas in Communications, IEEE Journal on*, 29(9):1872–1892, 2011.

[13] P. Grabowicz, J. Ramasco, E. Moro, J. Pujol, and V. Eguiluz. Social features of online networks: The strength of intermediary ties in online social media. *PLoS ONE*, 7(1):e29358, 2012.

[14] M. Granovetter. The strength of weak ties. *American journal of sociology*, pages 1360–1380, 1973.

[15] F. Heider. *The psychology of interpersonal relations*. Lawrence Erlbaum, 1982.

[16] M. Kurant, A. Markopoulou, and P. Thiran. Towards unbiased bfs sampling. *Selected Areas in Communications, IEEE Journal on*, 29(9):1799–1809, 2011.

[17] M. McPherson, L. Smith-Lovin, and J. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444, 2001.

[18] P. Mucha, T. Richardson, K. Macon, M. Porter, and J. Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876, 2010.

[19] T. Newcomb. *The acquaintance process*. Holt, Rinehart and Winston, 1961.

[20] M. Newman. Power laws, pareto distributions and zipf's law. *Contemporary physics*, 46(5):323–351, 2005.

[21] M. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):26113, 2004.

[22] J. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A. Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332, 2007.

[23] T. Opsahl, V. Colizza, P. Panzarasa, and J. Ramasco. Prominence and control: The weighted rich-club effect. *Physical review letters*, 101(16):168702, 2008.

[24] S. Pajevic and D. Plenz. The organization of strong links in complex networks. *Arxiv preprint arXiv:1109.2577*, 2011.

[25] A. Petróczi, T. Nepusz, and F. Bazsó. Measuring tie-strength in virtual social networks. *Connections*, 27(2):39–52, 2006.

[26] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow. The anatomy of the facebook social graph. *Arxiv preprint arXiv:1111.4503*, 2011.

[27] R. Xiang, J. Neville, and M. Rogati. Modeling relationship strength in online social networks. In *Proc. 19th international conference on World wide web*, pages 981–990. ACM, 2010.

[28] J. Zhao, J. Wu, and K. Xu. Weak ties: Subtle role of information diffusion in online social networks. *Physical Review E*, 82(1):016105, 2010.

[29] S. Zhou and R. Mondragón. The rich-club phenomenon in the internet topology. *Communications Letters, IEEE*, 8(3):180–182, 2004.