SCS M&S Magazine. Vol. IV. Issue 3.
M. Baldoni, C. Baroglio, and A. Garro (Guest eds.)

OPTIMIZING SOCIAL GROUP HOMOGENEITY IN ONLINE COMMUNITIES

Pasquale De Meo
Department. of Ancient and Modern Civilizations
University of Messina
98166 Messina, ITALY

Domenico Rosaci
Department. DIIES
University Mediterranea of Reggio Calabria
Loc. Feo di Vito
89122 Reggio Calabria, ITALY

Giuseppe M.L. Sarnè
Department. DICEAM
University Mediterranea of Reggio Calabria
Loc. Feo di Vito
89122 Reggio Calabria, ITALY

ABSTRACT

The optimization of homogeneity in social interest groups in Online Social Networks involves their formation and evolution and it is based on both users' preferences and choices of the administrators' groups. To consider the satisfaction of group members, also taking into account their similarity, the *homogeneity* of a social group it is often regarded as a basilar condition. In this paper we introduce a group homogeneity measure based on users' behavioral information. Such a measure is optimized by a novel algorithm that by operating in a fully-distributed multi-agent framework performs users and groups profiles matching. Such an approach produces clear and significant advantages in groups formation in an online social network context as it is shown by some experimental campaigns performed on simulated social networks.

1 INTRODUCTION

Currently Online Social Networks (OSNs) such as Facebook (http://www.facebook.com), Google+ (http://plus.google.com) and Twitter (http://twitter.com) every day are increased both in complexity (Catanese et al. 2012) and in scale and content (Lehmann 2012) and this is a consequence of their increased social impact (Metaxas and Mustafaraj 2012). In such a scenario social *groups*, i.e. sub-networks of users sharing common interests, play a relevant role (Buccafurri et al. 2004, Gauch et al. 1997, Messina et al. 2013, Rosaci and Sarnè 2014).

Some authors investigated on the relationships occurring between users and groups belonging to OSNs (Baatarjav, Phithakkitnukoon and Dantu 2007, Hui and Buchegger 2009, Kim et al. 2010). In particular, Hui and Buchegger (2009) analyzed four OSNs in order to compute the probability that a user joins a group; Baatarjav, Phithakkitnukoon and Dantu (2007) with Kim et al. (2010) studied the problems of choosing which group to join with for a single user and for a group of users, respectively. However, to the best of our knowledge, no study investigated on the evolution of an OSN group as a problem of matching between users and groups profiles.

Differently from the concept of *social profile*, already considered in the context of virtual communities (Lampe, Ellison, and Steinfield 2007), that of *group profile* is rather novel. The concept of group profile is useful for solving the problem of suggesting a user the groups he/she could join with in order to improve his/her satisfaction level. Usually, a group might be considered (*i*) as a set of nodes (i.e., users) more densely connected among each other than to the others (i.e., the group, formation is viewed as a

graph clustering problem (De Meo et al. 2013); or, (*ii*) as a community of people sharing similar interests (Currarini, Jackson and Pin 2009).
Moreover, the notion of *group homogeneity* is often related to that of *Similarity*. Indeed, in presence of a high similarity/inter-connectivity among group participant, according to both structural and semantic dimensions, an OSN group can be considered as homogeneous and this implies a better satisfaction among its users (Lewis, Gonzalez and Kaufman 2012). Therefore, when we assume *homogeneity* as related to users' satisfaction, we can suppose that other behavioral characteristics of members and groups should be considered as important characteristics (Centola 2010). For example, in virtual communities, users often have multiple interests; groups define common rules, considered as accepted behaviors, and exhibit several communication styles and implement several facilities for sharing media content.

In this paper, a novel measure of group homogeneity, based on users similarity and the other users' features cited above, is defined. By means such a definition, we provide an algorithm to match the individual users' profiles with those of groups. In such a way, we can find the matching between users and groups capable of improving the homogeneity of the social groups. More in detail:

- The notion of group profile in the context of OSNs is introduced. Coherently with the definition of a user profile, that of group profile considers a set of categories of interests, common rules, behaviors, communication styles and facilities for sharing media content.
- Each OSN group is associated with a group agent (De Meo et al. 2011) which create, manage and update the group profile defined above. Similarly, a user agent is associated with each OSN user.
- We present a distributed agent platform to handle group formation (Palopoli, Rosaci and Sarnè 2013, Rosaci and Sarnè 2013). The agents are capable to automatically and dynamically compute a matching between user and group profiles in a distributed manner. The user agent are provided with a matching algorithm, named Group Homogeneity Maximization (GHM), exploiting a homogeneity measure between user and group profiles that we introduced. By means of such a measure, the algorithm provides to determine the group profiles best matching user ones.
- The GHM algorithm will be executed to improve the intra-group homogeneity as follows:
  - the user agent submits some requests for joining with the best groups;
  - each group agent accepts only those requests sent by agents having their profiles matching with the group profile.
- An experimental evaluation of the GHM matching algorithm, performed on a set of simulated users and groups, doubtless has shown the advantages of our proposal.

It is important to point out that the effectiveness of the algorithm strictly depends on both the two tasks performed on the user-side and group-side, respectively. Indeed, the algorithm guarantees that the individual user joins only with groups whose group profiles are similar with his own profile, but also assures that each group accepts as new member only users whose profiles have a sufficient similarity with the group profile. In a real scenario, we imagine that the two tasks will be implemented by software agents assisting both users and group administrators, where the two types of agents behave as personal assistants for their human counterparts.
The choice of introducing an agent-oriented approach is motivated by the need of avoiding a centralized solution, which would imply significant limitation in efficiency. The multi-agent architecture that we propose allows to distribute the computation of the mutual similarities on the whole social network, by assigning to each actor (either user or group) the appropriate task. Moreover, each software agent execute its task autonomously and proactively, without involving the users in annoying interactions.
In the following, the paper is organized as follow. In Section 2 we introduces the General Guidelines; Section 3 presents the GMH algorithm and in Section 4 are described the experiments we performed to evaluate our method and its advantages and limitations. Related work are discussed in Section 5 and, finally, in Section 6 some conclusions are drawn.

2 GENERAL GUIDELINES

In the proposed scenario, let to consider an OSN with the sets of its users and groups, that we denoted by S, U and G, respectively. More in detail, in S, each group of users $g \in G$ represents a subset of U such that $g \subseteq U$, $\forall g \in G$. Moreover, a multi-agent system is associated with S (see Figure 1), such that: (*i*) each user u is assisted by his/her personal agent a_u in performing the activities of participation to groups; and, (*ii*) each group g is supported by an administrator agent a_g in the task of managing all the received agent requests to join with the group.

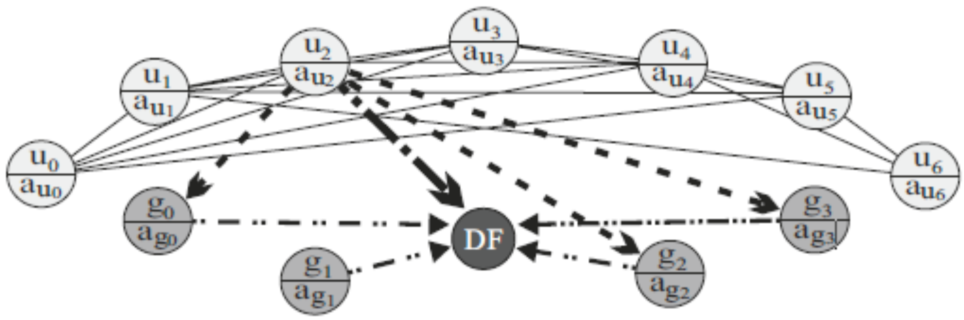

Figure 1. The multi-agent architecture.

2.1 The Agents Knowledge

The knowledge that each agent a_u (resp., a_g) has about the interests and preferences of its user u (resp., group g), is represented by means of a profile p_u (resp., p_g) associated with it. Such a profile stores preference and behavioral information of the user u (resp., the users of g) in four sections called interests, access preference, behaviors and friends that respectively store data on topics of interest, mode to access groups, ways of performing activities and friends. The profile of a user u (resp., a group g) is represented by a 4-tuple $\langle I_u , A_u , B_u , F_u\rangle$ (resp., $\langle I_g , A_g , B_g , F_g\rangle$), where each component describes the properties of u (resp., g).

Let C be the set of all categories considered in the OSN, where each element $c \in C$ is an identifier devoted to represent a given category (e.g. *music*, *sport*, etc.). Each OSN user u (resp., group g) deals with some categories belonging to C where I_u (resp., I_g) denotes a mapping that returns a real value $I_u(c)$ (resp., $I_g(c)$), ranging in [0..1], for each category $c \in C$. This value represents the interest level of the user u (resp., the users belonging to the group g) with respect to discussions and multimedia content dealing with c. In computing the values of this mapping are taken into account the actual behavior of u (resp., of the users of g), see Section 2.2 for the details.

The access mode property is that property representing the policy regulating the access to a group (that is described by an identifier, e.g. *open*, *closed*, *secret*, etc.) preferred by u (resp., set by the administrator of the group g) and denoted by A_u (resp., A_g).

The property B_u describes the types of behavior adopted (resp., required) by u in his/her OSN activities, for example "publishing posts shorter than 500 characters". Let $b \in B$ a behavior adoptable by user u (resp., admitted in the group g) and described by a Boolean variable set to *true* if b is adopted (resp., tolerated), or *false* otherwise, and let B be the set of possible behaviors associated with the OSN (e.g., $B = \{b_1 , b_2 ,\cdots, b_n\}$). Therefore, let B_u (resp., B_g) be a mapping that, for each $b \in B$, returns a Boolean value $B_u(b)$ (resp., $B_g(b)$), where $B_u(b_i) =$ *true* means that such behavior is adopted by u (resp., tolerated in g), or *false* otherwise.

Finally, the property F_u (resp., F_g) represents the set of all users that are friends of u (resp., that at least have a friend among the members belonging to the group g).

2.2 The Agents Tasks

After that u (resp., a user affiliated to g) performs an action involving an information stored in its profile in turn the agent a_u (resp., a_g) automatically performs the task to update the profile p_u (resp., p_g) of its user u (resp., group g).

In particular, every time u deals with a category c, the associated value $I_u(c)$ is updated as the weighted mean between its previous value and the new contribution to $I_u(c) = \alpha \cdot I_u(c) + (1 - \alpha) \cdot \delta$. In detail, α and δ are real values arbitrarily autonomously set by u in [0..1], where δ is the increment to give to the u's interest in c due to his/her action, while α weights the two components of $I_u(c)$. Similarly, every time the $I_u(c)$ value of any user $u \in g$ changes, the $I_g(c)$ value of a group g is updated by the agent a_g as the mean of all the $I_u(c)$ values $\forall c \in g$.

For each action performed by the user u, for instance like to publish a post, its agent a_u sets in B_u the Boolean values of the associated variables. Similarly, the agent a_g updates the variables contained in B_g every time the administrator of g changes the associated rules. Besides, when u (resp., the administrator of g) modifies his/her preferences about the access mode, the associated agent provides to update A_u (resp., A_g). Also, when u (resp., a user of g) modifies his/her friends list, the associated agent updates F_u (resp., F_g). Note that a_g computes F_g as the union of the sets F_u of all the users belonging to g.

Periodically, the agent a_u (resp., a_g) provides to execute the user (resp., group) agent task described above, to contribute to the group matching activity of the OSN.

To perform the above tasks, the agents can reciprocally interact, send and receive messages thanks to a Directory Facilitator agent (*DF*), associated with the OSN. This agent provides an indexing service. In particular, the *DF* stores the names of each user and group belonging to the OSN and those of their agents. Note that the *DF* is the only one centralized component in our scenario, while the GHM matching algorithm results to be completely distributed on the whole agent framework.

2.3 Definition of Homogeneity

To represent the potential attitude of the user u to stay in the same group with the user v (resp., to stay in the group g), we define the homogeneity between two users u and v (resp., a user u and a group g) as a measure representing how much u and v (resp., u and g) are similar (or, different) with respect to the considered properties I, A, B and F.

Our definition of homogeneity relies on the concept of *homophily*, i.e., the tendency of individuals to aggregate based on shared interests (McPherson, Smith-Lovin and Cook, 2001). Therefore, we suppose that if two individuals share the same interests, they are more likely to join and stay in the same group. In addition, we claim that social relationship play a relevant role in pushing a user to join a group: to this purpose we cite the study proposed in (Backstrom et al., 2006) in which the authors showed that the probability that a user joins a group increases (in a sublinear fashion) when the fraction of his/her friends who already joined the group increases too.

We therefore suggest to separately considering user similarities according to several dimensions like interests, behaviors and social relationships. We then compute the contribution to homogeneity carried out by each of these dimensions and we combine single contributions to get a global one. There are, in principle, several ways to aggregate single contributions to compute homogeneity. In this paper we focused on the simplest one, i.e., we choose to define the homogeneity as the weighted mean of each of the contributions cited above. We left as future study the usage and experimental comparison of further (and more sophisticated) aggregation techniques.

The homogeneity $h_{u,v}$ between the users' profiles of u and v is defined as a weighted mean of the contributions c_I, c_A, c_B and c_F respectively associated with the properties I, A, B and F and measuring how much the values of each property in p_u and p_v are similar. To this purpose:

- c_I is the average of the differences (in the absolute value) of the interests values of u and v for all the categories present in the social network, that is $c_I = \sum_{c \in C} |I_u(c) - I_v(c)| / |C|$.
- c_A is set to 0 or 1 if A_u is equal or not equal to A_v.
- c_B is the average of all the differences between the Boolean variables stored in B_u and B_v, where this difference is set to 0 or 1 if the two corresponding variables are equal or different.
- c_F is computed as the percentage of common friends of u and v, with respect to the total number of friends of u or v as $c_F = |F_U \cap F_v| / |F_U \cup F_v|$. Note that, to make them comparable, the contributions are normalized in [0..1].

The homogeneity $h_{u,v}$ is then computed as:

$$h_{u,v} = \frac{w_I \cdot c_I + w_A \cdot c_A + w_B \cdot c_B + w_F \cdot c_F}{w_I + w_A + w_B + w_F} \tag{1}$$

Similarly, homogeneity $h_{u,g}$ between a user u and a group g is simply computed as $h_{u,v}$ by substituting user v with the group g.

3 THE GHM ALGORITHM

The GHM algorithm is an activity globally distributed in the framework and periodically executed by each user agent a_u and each group agent a_g, where we call epoch every time the task is executed and T the (constant) period between two consecutive epochs.

3.1 The User Agent Task

Let X be the set of the n groups which u is affiliated to, where $n \leq n_{MAX}$ and n_{MAX} is the maximum number of groups a user can join with. We suppose that a_u stores into a cache the profile p_g of each group $g \in X$, that it contacted in the past, together with the date $date_g$ of its acquisition. Moreover, let m be the number of group agents that is contacted by a_u at each epoch. In such a context, a_u behaves as follows (see Figure 2):

- From the *DF* repository a_u provides to randomly select a set Y of m groups so that $X \cap Y = \{0\}$ and let $Z = X \cup Y$ the set consisting of all the groups present in X or in Y.
- For each group $g \in Z$ such that $date_g > \psi$ (where ψ is a fixed system threshold), u sends a message to the agent a_g for asking the profile p_g associated with g (cf. Action 1 of Figure 2).
- For each received p_g (cf. Action 2 of Figure 2), u computes a homogeneity measure $h_{u,g}$ between his/her profile and that of the group g (cf. Action 3 of Figure 2).
- The groups belonging to Z and having the highest homogeneity values such that $h_{u,g} > \tau$, where τ is a real value ranging in [0..1], are inserted by a_u in the set of good candidates, named *GOOD*, to join with (up to a maximum of n_{MAX} groups). For each group $g \in GOOD$ if $g \notin X$, then a_u sends a join request and the profile p_u of u to a_g (cf. Action 4 of Figure 2). Otherwise, if $g \in X$ but $g \notin GOOD$, then a_u deletes u from g.

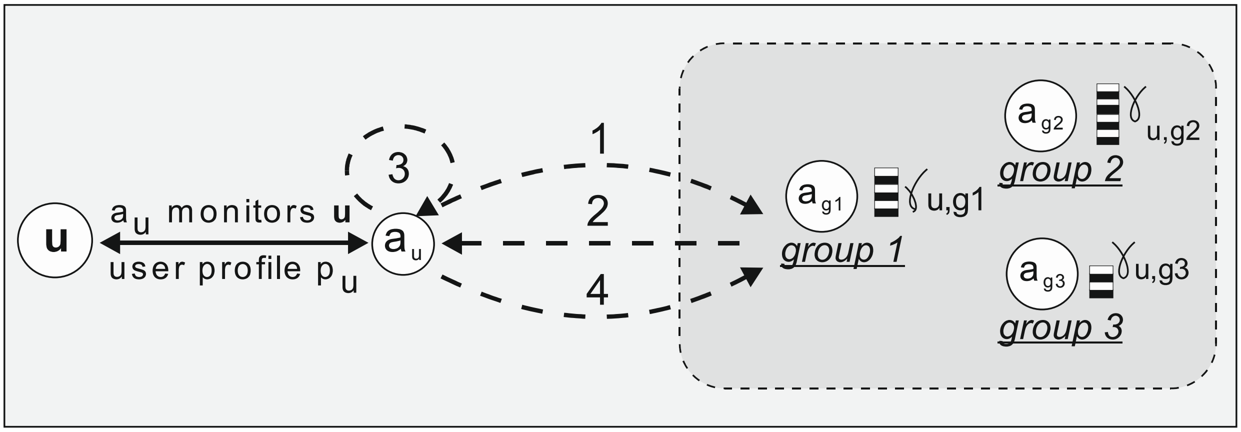

Figure 2. User agent task schema.

From a computational standpoint, observe that the most computationally intensive parts of the GHM algorithm consists in the computation of the set Z and in the identification of the groups having the largest homogeneity.
To this purpose, (Ding and Konig, 2011) recently introduced a linear data structure allowing to compute the intersection of two or more sets in sublinear time with respect to the total number of available elements in the input set. Analogously we can use ad-hoc data structures to keep the set of available groups sorted on the basis of their homogeneity.

3.2 The Group Agent Task

Let K be the set of the k users affiliated to the group g, where $k \leq k_{MAX}$, being k_{MAX} the maximum number of members allowed by the administrator of g. Suppose that into its cache a_g stores the profiles of the users $u \in K$ obtained in the past along with the date $date_u$ of their acquisition. When a_g receives a join request by a user agent u (along with u's profile p_u), it behaves as follows (see Figure 3):

- For each user $u \in K$ such that $date_u > \eta$ (i.e., a fixed system threshold), it sends a message to the agent a_u to require the profile p_u associated with u (cf Action 1 of Figure 3).
- When a_g receives the required users' profiles (cf. Action 2 of Figure 3), it computes the homogeneity measure $h_{g,u}$ between the profile of each user $u \in K \cap \{r\}$ and the profile of the group g (cf. Action 3 of Figure 3).
- The user u having the highest homogeneity values such that $h_{g,u} > \pi$, where π is a real value ranging in [0..1], is inserted by a_g in the set of good candidates, named *GOOD*, to join with (up to a maximum of k_{MAX} users). If $u \in GOOD$, then a_g accepts its request to join with g (cf. Action 4, Figure 3). Moreover, if $u \in K$ but $u \notin GOOD$, then a_g deletes u from g.

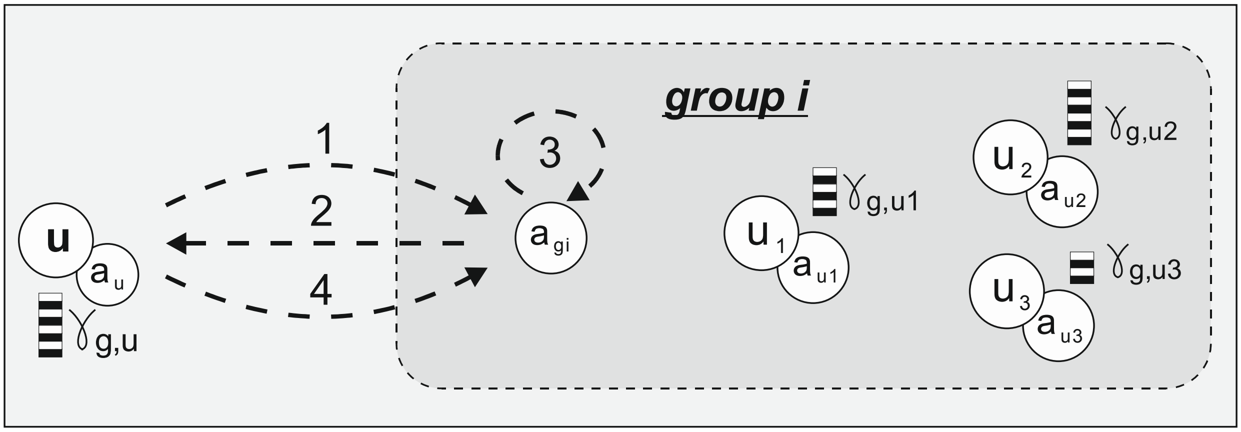

Figure 3. The group agent task schema.

4 EVALUATION

In order to evaluate the effectiveness of the GHM algorithm in obtaining an increasing in the homogeneity of the groups of an OSN, we have built a simulator, called GHM-Sim, capable of modeling all the required users and groups activities. The experiments involve a simulated OSN having 30.000 users and 100 groups, ad hoc generated by GHM-Sim, each one provided with a profile, having the structure described in Section 2. More in detail, the profile p_u of a user u is generated as follows:

- The values of $I_u(c)$ are randomly chosen from a uniform distribution in the interval [0..1];
- A_u is assigned the value open (resp., *closed* and *secret*) with a probability of 0.7 (resp., 0.2, 0.1) to implement the variability of OSNs group access restrictions;
- B_u contains the values, randomly generated, of six Boolean variables representing in average the user's attitude to:
  - publish more than 1 post per day;
  - publish posts longer than 200 characters;
  - comment at least two posts of other users per day;
  - respond to comments associated with her posts;
  - leave at least 2 "Like" rates per day;
  - respond to the messages.
- The set of friends F_u are randomly generated by choosing in the set of the users.

Users are initially randomly assigned to at least 2 and at most 15 of the available groups. The properties I_g, A_g, B_g and F_g of the profile p_g of each group g are randomly generated. The values of the parameters introduced in Section 3 are shown in Table 1. We also limit to:

- 250 are the users who can join a given group;
- 15 are the groups that a user can be joined with;
- 5 is the maximum number of requests that a user can send in each epoch to new groups.

Table 1: The parameters values used in the GHM-Sim simulator

Parameter	Value
τ	250
π	15
K_{MAX}	5
N_{MAX}	0.4
N_{REQ}	0.4

To measure the internal *homogeneity* of a group g we use the average homogeneity AH_g, derived by (Pearson et al. 2004), computed as $\sum_{x,y\in g,x\neq y} h_{x,y} / |g|$, while to measure the *global homogeneity* of the OSN groups we compute the *mean average homogeneity MAH* and the *standard deviation average homogeneity DAH* of all the AH_g, defined as

$$MAH = \frac{\sum_{g\in G} AH_g}{|G|} \tag{2}$$

$$DAH = \sqrt{\frac{\sum_{g\in G} \left(AH_g - MAH\right)^2}{|G|}} \tag{3}$$

In the simulations, the initial values for the above measures were MAH = 0.266 and DAH = 0.0011, denoting a very low homogeneity, due to the random generation. Applying the GHM algorithm, we have simulated 15 epochs of execution per user. We can observe that the GHM algorithm quickly converges after few iterations (see Figure 4). The experimental results show that the GHM algorithm increases the homogeneity in OSN groups of about 14 percent on average, with respect to a random assignment of users to groups, achieving a stable configuration (e.g., MAH = 0.320 and DAH = 0.0052) after about 10 epochs. It is reasonable to suppose that the GHM algorithm, when applied to real OSNs, should lead to concrete benefits in terms of homogeneity.

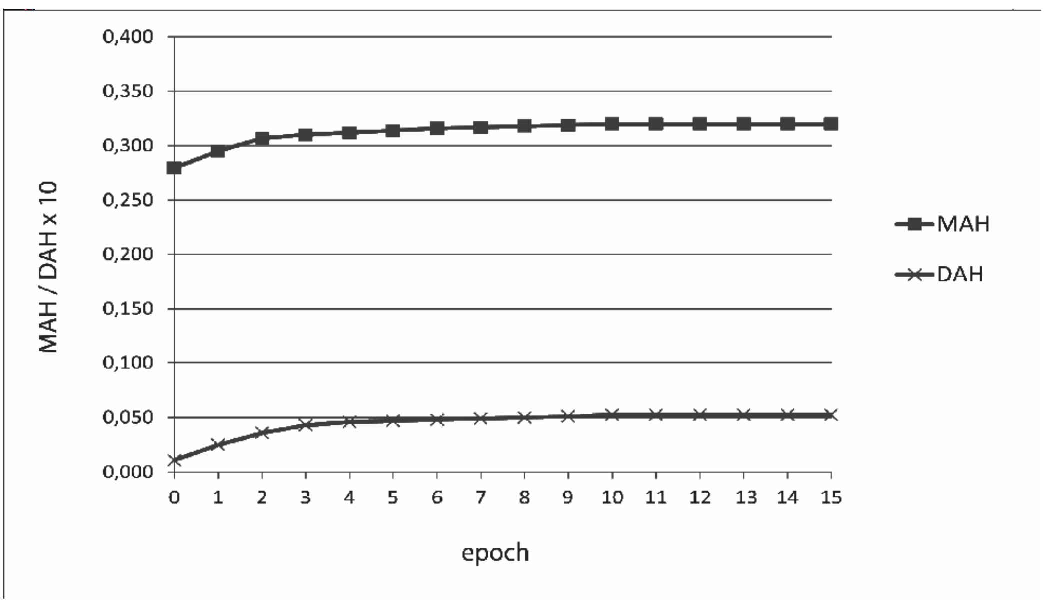

Figure 4. Variation of *MAH* and *DAH* (x10) vs epochs obtained with the GHM-comp and GHM-diff algorithms, for a SN with 30.000 users and 100 groups.

As a final experiment, we studied the computational performance of our GHM algorithm. All the experiments were carried out on a PC equipped with an Intel i7 Quadcore with 8 Gb of RAM, 500 Gb of HD and Windows 8 operating system. Experiments show that the GHM algorithm converges in at most 3 or 4 epochs, i.e., the association among users and groups do not significantly change after 5 execution of the algorithm itself. As observed in Section 3.1, the execution of each epoch is fast if we adopt suitable data structures to represent groups and their homogeneity values. Therefore, we can conclude that our algorithm is scalable and it can manage large social network instances.

5 RELATED WORK

In this section, we compare our work with research done in the field of *group modeling* and in the *computation of user similarities in Online Social Networks.*

With respect to the field of *Group Modeling*, in the latest years there was a growing interest in the task of producing recommendations for the members of a group. This requires to generate a model capable of representing the group interests and needs and it fits quite well the concept of group profile provided in our paper (Amer-Yahia et al 2009, Baltrunas et al. 2010).

The task of creating a profile to associate with a group is often called *group modeling*. Most of the existing approaches to group modeling starts by aggregating the preferences of group members to produce a global group profile (Amer-Yahia et al 2009, Gorla et al. 2013). These preferences are usually encoded as numerical scores (ranging in a discrete interval like [1,..,5]).

We focus on two specific strategies, namely:

- *Average*, in which the score of an item i for a group G is equal to the average of the scores assigned by the members of G to i;
- *Least Misery*, i.e., the score of an item i for the members of a group G is equal to the least score the members of G assigned to i.

Observe that none of the two strategies dominates the other one: for instance, if we would adopt the Average strategy, then an item which is highly rated will be recommended to all the members of G, even if some members of G may totally dislike it. In contrast, if we would adopt the Least Misery strategy, then an item will automatically get a low score if there exist at least a user who dislikes it and all the remaining users like it.

Most of the approaches to group modeling assume that user's preferences are independent of the fact the a user has decided to join with a group or not: therefore, if a user alone likes (or dislikes) an item he/she will continue liking (or disliking) it if he/she decides to join with a group. Such an assumption could be harshly restrictive: in fact, due to social influence phenomena, users can change their original opinion prior (or after) joining a group. Recently, Gorla et al. (2013) described a probabilistic framework capable of modeling the preferences that arise in an individual when he/she joins with a group.

Our approach differs from the approaches cited above because it provides a rich framework modeling user interests and friendship relationships, as well as information describing user behaviors. Our framework allows for modeling the policies followed to access groups and previous user behaviors like the posts he/she generated or if he/she liked/disliked an item. A further difference is that the management of both group and user profiles is carried out by means of a multi-agent architecture and agents are in charge of updating user and group profiles as well as of finding groups a user could join.

In *Computing user similarities* in the context of online social networks, the problem of detecting whether two users are similar has been extensively studied.

We may classify approaches to computing user similarities into two main categories, namely:

- Approaches relying on social relationships;
- Approaches based on the analysis of social activities.

Our definition of homogeneity is *hybrid*, in the sense that it combines both the features of approaches relying on social relationships (because we take user friendship into account) and the features of approaches based on the analysis of social activities (because we handle, for instance, the type of behaviours of a user).

As claimed before, a first category of approaches relies on social relationships existing among users. In many cases, this information is instrumental in producing suggestions (e.g., friendship relationships or affiliation to new communities).

In particular, the approach of (Spertus, Saham and Buyukkokten 2005) analyses the affiliation of users to multiple virtual communities; users are advised if it is/is not convenient to join a community. To this purpose, their approach considers *Orkut*, a big social network, as reference scenario and experimentally compares the effectiveness of 6 similarity measures (e.g., *tf-idf* coefficient or parameters coming from Information Theory). The approach of (Groh and Ehmig 2007) suggests to use the friendship lists to identify resources relevant to users. In particular, the approach of (Groh and Ehmig 2007) handles the friendship list of a user u and the ratings of the users of these lists assigned to an object o to predict the rating that u would assign to o.

Approaches relying on social relationships are able to achieve a high level of accuracy in producing recommendations (see (Groh and Ehmig 2007) for an experimental analysis). In addition, these approaches are less plagued by problems like *cold start*. The effectiveness of these approaches, however, crucially depends on the number of social relationships created by users. In fact, if a user is involved in few friendship relationships, the information at disposal are sparse and, then, the quality of suggestions will be inevitably poor.

Our approach merges the analysis of social relationship with behavioral information. This kind of information is a reliable indicator to assess whether two users are similar or not even if they do not know directly.

Approaches based on the analysis of social activities rely on the idea that if two users participate to the same activities, then a form of similarity between them can be envisaged. In particular, information associated with a user contributes to form a profile capable of describing her preferences and needs. The similarity between two users is then computed by taking into account the similarity of their profiles.

In (de Gemmis et al. 2008) the authors consider the tags applied by users to classify resources and provide a generative probabilistic model to build their profile. (Pazzani and Billsus 1997) use a number of machine learning techniques (like Bayesian classifiers or decision trees) to analyze Web pages accessed by the user and build her profile. In (De Meo, Quattrone and Ursino 2010) the authors propose to analyze semantic relationships between tags applied by users to classify folksonomy resources and use these tags to enrich user profiles.

Our approach, like those described in this section, considers the activities that the users of a social network can carry out. We focus on the activities a user can be involved to as well as the mode (open, closed, secret) and the type of behavior adopted by a user whereas the approaches illustrated in this section, rely on activities like *tagging* (de Gemmis et al. 2008, De Meo, Quattrone and Ursino 2010) or *browsing* (Pazzani and Billsus 1997). The analysis of user activities provides useful elements to generate accurate and complete profiles.

6 CONCLUSIONS

The problem of dynamically increasing the intra-group homogeneity is emerging as a key issue in the OSN research field. The introduction of high-structured user profiles, the large dimensions of current OSNs and the increasing number of groups require to face efficiency and scalability issues.

In this paper, we presented the Group Homogeneity Maximization algorithm that allows a set of software agents, associated with the OSN user profiles, to dynamically and autonomously manage the evolution of the groups, detecting for each user the best groups to join with based on the measures of homogeneity. The agents associated with the group administrators accept only those users having a profile compatible with that of the group. Our experiments on simulated social network data clearly show that the execution of the matching algorithm increases the internal homogeneity of the groups composing the social network, bringing about 15% of improvement with respect to the baseline.

We argue that the improvement of the internal homogeneity introduced by the algorithm should corresponds to an increment of the users satisfaction, consequent to the fact the each user should consider more profitable for him to stay in groups where the members have similar interests and behaviors. However, such a supposition need to be confirmed by experiments on real data. The results of our simulations are limited to only validate the effectiveness of our approach in incrementing the homogeneity.

In order to obtain more accurate results, in our ongoing research we are considering to combine the homogeneity measure with a new measure taking into account the trustworthiness of the users. Indeed, in virtual communities, interacting users reciprocally measure the trustworthiness of their counterparts to decide if these are reliable interlocutors or not. To this aim, we are planning a specific experimental session on real OSN data to evaluate our approaches.

REFERENCES

Amer-Yahia, S., S. Roy, A. Chawlat, G. Das, and C. Yu. 2009. "Group Recommendation: Semantics and Efficiency". In *Proceedings of VLDB Endowment*, 2(1):754–765.

Baatarjav, E., S. Phithakkitnukoon, and R. Dantu. 2008. "Group Recommendation System for Facebook". In *On the Move to Meaningful Internet Systems: OTM 2008 Workshop*. 211–219. Springer.

Backstrom, L., Huttenlocher, D.P., Kleinberg, J.M. and Lan, X, 2006. "Group formation in large social networks: membership, growth, and evolution." *Proc. of the International Conference on Knowledge Discovery from Data (KDD 2006)*, 44-54

Baltrunas, L., T. Makcinskas, and F. Ricci. 2010. "Group Recommendations with Rank Aggregation and Collaborative Filtering". In *Proceedings of ACM Conference on Recommender Systems 2010*. 119–126. ACM Press.

Buccafurri, F., D. Rosaci, G. M. L. Sarne, and L. Palopoli. 2004. "Modeling Cooperation in Multi-agent Communities". *Cognitive Systems Research*. 5(3):171–190.

Catanese, S., P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, 2012. "Extraction and Analysis of Facebook Friendship Relations". *Computational Social Networks.* Edited by Abraham, A. 291-324. Springer

Centola, D. 2010. "The Spread of Behavior in an Online Social Network Experiment". *Science* 329:1194–1197.

Currarini, S., M.O. Jackson, and P. Pin. 2009. "An Economic Model of Friendship: Homophily, Minorities, and Segregation". *Econometrica* 77(4):1003–1045.

de Gemmis, M., P. Lops, G. Semeraro and P. Basile. 2008 "Integrating tags in a semantic content-based recommender". In *Proceedings of the ACM Conference on Recommender* Systems (RecSys '08). 163-170. ACM Press.

De Meo, P., G. Quattrone and D. Ursino. 2010. "A query expansion and user profile enrichment approach to improve the performance of recommender systems operating on a folksonomy". *Journal of User Modelling and User Adapted Interactions*, 20(1):41-86.

De Meo, P., A. Nocera, D. Rosaci, and D. Ursino. 2011. "Recommendation of Reliable Users, Social Networks and High-Quality resources in a Social Internetworking System". *AI Communications* 24(1):31–50.

De Meo, P., E. Ferrara, G. Fiumara, and A. Provetti. 2014. "Mixing Local and Global Information for Community Detection in Large Networks". *Journal of Computer and System Sciences,* 80(1): 72-87

Ding, B., and König, C., 2011. "Fast set intersection in memory", *Proceedings of the VLDB Endowment,* 4(4): 255-266.

Firan, C.S., W. Nejdl and R. Paiu. 2007. "The benefit of using tag-based profiles". In: *Proceedings of the Latin American Web Congress (LA-Web 2007)*, 32–41. IEEE Computer Society.

Gauch, S., M. Speretta, A. Chandramouli, and A. Micarelli. 2007. "User Profiles for Personalized Information Access". In *The Adaptive Web*, volume 4321 of *LNCS*, 54–89. Springer.

Gorla, J., N. Lathia, S. Robertson, and J. Wang. 2013. "Probabilistic Group Recommendation via Information Matching". In *Proceedings of the International World Wide Web Conference (WWW '13)*. 495–504. ACM Press.

Groh, G. and C. Ehmig. 2007. "Recommendations in taste related domains: collaborative filtering vs. social filtering". In *Proceedings of the International ACM conference on Supporting Group Work (GROUP '07).* 127-136, ACM Press.

Hui, P. and S. Buchegger. 2009. "Groupthink and Peer Pressure: Social Influence in Online Social Network Groups". In *Proceedings of the International Conference on Advances on Social Network Analysis and Mining*. 53–59. IEEE.

Kim, J., H. Kim, H. Oh, and Y. Ryu. 2010. "A Group Recommendation System for Online Communities". *International Journal of Information Management*. 30(3):212–219.

Lampe, C.A., N. Ellison, and C. Steinfield. 2007. "A Familiar Face (Book): Profile Elements as Signals in an Online Social Network". In *Proceedings of the SIGCHI Conference*. 435–444. ACM Press.

Lehmann, J., B. Gonc¸alves, J. Ramasco, and C. Cattuto. 2012. "Dynamical Classes of Collective Attention in Twitter". In *Proceedings of the 21st International Conference on World Wide Web.*251–260.

Lewis, K., M. Gonzalez, and J. Kaufman., 2012. "Social Selection and Peer Influence in an Online Social network". *Proceedings of National Academy of Sciences*, 109(1):68–72.

McPherson, M., Smith-Lovin, L., and Cook, J. M., 2001. "Birds of a Feather: Homophily in Social Networks". Annual Review of Sociology. 27:415–444.

Messina, F., G. Pappalardo, D. Rosaci, C. Santoro, and G. M. L. Sarnè. 2013. "Hyson: A Distributed Agent-based Protocol for Group Formation in Online Social Networks". In *Multiagent System Technologies*. Series LNCS. Volume 8076. pp. 320–333. .

Metaxas, P.T., and E. Mustafaraj. 2012. "Social Media and the Elections". *Science* 338(6106):472–473.

Palopoli, L., D. Rosaci, and G. M. L. Sarnè. 2013. "Introducing Specialization in e-Commerce Recommender Systems". *Concurrent Engineering: Research and Applications*, 21(3):187–196.

Pazzani, M., and D. Billsus. Learning and Revising User Profiles: The Identification of Interesting Web Sites. *Machine Learning*. 27(3):313-331, 1997.

Pearson, R., T. Zylkin, J. Schwaber, and G. Gonye. 2004." Quantitative Evaluation of Clustering Results Using Computational Negative Controls". In *Proceedings of SIAM International Conference on Data Mining.* 188–199.

Rosaci, D., and G.M.L. Sarnè. 2013. "Recommending Multimedia Web Services in a Multi-device Environment". *Information Systems* 38(2):198–212.

Rosaci, D., and G.M.L. Sarnè. 2014. "Matching Users with Groups in Social Networks". *Intelligent Distributed Computing VII*, Edited by F. Zavoral, J.J. Jung, and C. Badica. Volume 511 of *Studies in Computational Intelligence*. 45–54. Springer.

E. Spertus, M. Sahami, and O. Buyukkokten. 2005. "Evaluating Similarity Measures: a Large-scale Study in the Orkut Social Network". In *Proceedings of ACM International Conference on Knowledge Discovery and Data Mining*. 678–684. ACM Press.

AUTHOR BIOGRAPHIES

PASQUALE DE MEO is an assistant professor of Computer Science at the University of Messina, Italy. His main research interests are in the area of Social Web, user modelling and recommender systems. His email address is pdemeo@unime.it.

DOMENICO ROSACI is assistant professor of Computer Science at University of Reggio Calabria, Italy. His research focuses on Distributed Artificial Intelligence, mainly investigating social aspects of multi-agent communities, recommender systems and social networks. His email address is domenico.rosaci@unirc.it.

GIUSEPPE M. L. SARNE' is assistant professor of Computer Science at University of Reggio Calabria, Italy. His main research interests are in Distributed Artificial Intelligence and in particular on social aspects of multi-agent, recommender, social networks and e-Commerce systems. His email address is sarne@unirc.it.