Improving Agent Group Homogeneity Over Time

Pasquale De Meo*, Fabrizio Messina†, Domenico Rosaci‡, Giuseppe M. L. Sarné ‡
*University of Messina, Italy, pdemeo@unime.it
†University of Catania, Italy, messina@dmi.unict.it
‡University of Reggio Calabria, Italy, {domenico.rosaci,sarne}@unirc.it

***Abstract*—In social communities the composition of thematic groups varies over time due to changes occurring in users' behaviors. To study the time evolution of such a process, we design a conceptual framework exploiting a distributed algorithm driving group formation. The results of tests carried out on real data extracted by the social network CIAO, show as groups formed by combining similarity and trust measures are *i*) more time-stable, independently by the weight of the trust component, and *ii*) more time-homogeneous, independently by the presence of uncorrelated random agents' behaviors affecting the similarity component.**

***Index Terms*—Homogeneity, Similarity, Social Communities, Trust.**

I. INTRODUCTION

Many online communities include thematic *groups* [1] and many studies investigated on the users motivations to join with groups [2], as well as the impact of their growth [3], [4] and failure [1]. Important issues in forming groups require to an agent of selecting those groups able to satisfy its expectation [5]–[8] and, in a complementary way, the members of a group search to accept only new agents able to improve their utility. Many studies present algorithms effective in driving such group formation processes (e.g., by using diffusion processes [3], [4]) by analyzing *i*) the group evolution in terms of surviving or failure and *ii*) the reasons for which an agent will join with/left a group. Many of such studies assume that groups should be formed by like-minded members. To this aim, many measures exist to evaluate the similarity degree among the groups members (e.g., based on the number of affiliations to groups [7], member/group interests [5], individual preferences [6] and so on).

In such a context, the aptitude of a group to retain its members (i.e., *time stability*) [9], [10] is a major factor for its survival or extinction. The evolution of these groups is, in the most part of the cases, driven only by the mutual similarity criterion [11], although it has evident limitations. Indeed, the problem of recommending groups to a potential member usually relies only on the mutual similarity between the profile features of a candidate at the time t and the interests of a group. When such interests vary in a relatively short time frame then groups selected at the time t could not be the best choice at the time $t + \Delta t$.

Based on data of the social network CIAO [12], we verified as the mutual similarity alone does not ensure the group homogeneity over time. In particular, when uncorrelated (e.g., random) users' behavior aspects assume a relevant weight. Consequently, we investigated on improving the time-stability of the group homogeneity in terms of similarity. In this respect, recent studies on group formation processes consider *trust* [13]–[15] to increase the level of the group member's engagement over time and avoiding the group failure [16]. However, all the approaches reviewed in [16] do not face the problem of combining similarity with trust.

In [17]–[20] we proposed to integrate similarity and trust in a unique measure to form groups and finding those most suitable for joining with. In this paper, we extended this work to study how changes occurring in the similarity and trust measures impact on the groups formation over time. To test this approach, we used a new conceptual framework by means of which we verified, always by using the data of the social network CIAO, that groups formed by considering both similarity and trust have higher time-stable homogeneity, in terms of similarity, than groups formed by adopting the only similarity criterion. This result is valid also when both the weight assigned to the trust component is low or random (i.e., "uncorrelated") behavior components considered in computing the similarity have a relevant weight. We assume that this behavior is mainly due to the effect of trust in balancing potential incongruence of the similarity measures. Besides, we found as the function used to aggregate similarity and trust measures is not fundamental. Therefore, for a good trade-off between the need of producing accurate results and that of having an easy interpretation model, we aggregated similarity and trust by simply computing their weighted sum, as in [17].

The remaining of the paper is structured as follows: in Section II we present the related literature, in Section III the adopted reference scenario is introduced, while in Section IV deal with the similarity and trust measures computation. Sections V and VI illustrate the A2G algorithm and our conceptual framework. Section VII describes the experimental campaign, discuss its results and, finally, conclusions ends the paper.

II. RELATED WORK

In the group formation the *common identity and bond* theory [21] identifies as main mechanisms to join with a group (*i*) the strong personal ties and (*ii*)) the shared interests with other group members. Differently, in [2] the authors applied the community detection algorithms to identify clusters into OSNs and comparing their structural features with user-defined communities. A study on the group affiliation mechanisms is presented in [3], where in two different kind of networks the authors noted as the act to join with a group can be modeled in terms of new ideas spreading for both the networks.

Kairam *et al.* [4] compared the growth processes in which groups attract new members in *presence/absence* of social ties with *existing/any* group members. Their main finding is that more a group is highly clustered and more likely it grows for a diffusion process. It is in accord with [1], where 500,000 Facebook groups created in an 8-day period were monitored for 3 months after which the most part of groups were inactive. The analysis of [1] highlighted that group survival depends on the social capital brought by group founders and of their behavior. Differently from us, any of the papers above explored the homogeneity of such groups to verify if they result time-stable homogeneous in terms of similarity.

Homogeneous processes are defined as those whose parameters are time-stable with respect to some measure. Usual group formation processes do not assure that properties like to similarity, social identity and so on, will be time-stable. Conversely, some *affiliation recommendation* [7] processes suggest to an OSN user only time-stable groups. From a wide experimentation involving *Social Identity* and *Cohesion* theories on more Twitter datasets, the authors of [22] verified that groups based on the users' social identity and formed by users interested in a great variety of topics are less cohesive over time in presence of transient events.

A flexible framework in which group affiliation is treated as an event impacting on user's preferences is proposed and validated in [6], where a probabilistic framework provides to model the individual preferences when he/she joins with a group. In other words, it affiliates to a group those users maintaining a high similarity level into the group over time and keeping homogeneous the group under this point of view. In this scenario, we note as friendships and time-homogeneity of their relationships strictly depend by the mutual trust among individuals. Differently, in the literature the most part of the proposals to form time-stable groups consider it as a problem essentially involving some form of similarity among users.

Among the cited contributions, only [9] indirectly refers to trust, in the mean derived by the social theories [23], while the other consider some form of similarity as the unique criteria to form possible time-stable homogeneous groups.

III. The Reference Framework Scenario

Our framework involves a community $\mathcal{C} = \langle \mathcal{A}, \mathcal{G} \rangle$, where $\mathcal{A}$ and $\mathcal{G}$ are the sets of *agents* and *groups* active in $\mathcal{C}$. Let $\mathcal{I}$ a set of available *items*, each one belonging to a specific category x, belonging to the set of categories $\mathcal{X}$[1], that in $\mathcal{C}$ can be reviewed by each agent $a \in \mathcal{A}$ with an integer in the range $[0, 5]$. Moreover, let $\mathbf{r}_{a,i}$ be the generic *review* of an item $i \in \mathcal{I}$ released by a, which consists of: (*i*) a rating assigned to i by a; (*ii*) a category $x \in \mathcal{X}$ associated with i; (*iii*) a numerical score specifying the helpfulness[2] of $\mathbf{r}_{a,i}$; (*iv*) a timestamp. Finally, we denote by $\overline{\mathbf{r}}_a$ the set of reviews associated with a, called *review history*, and by $\mathcal{R}$ the set of all the review history in $\mathcal{C}$. In such a scenario, we assume to adopt a multi-agent platform where, each agent a_u assists a user u, whereas an administrator agent a_g assists a group g. The agents knowledge representation, the agent tasks and our definitions of similarity and trust will be introduced below.

[1]The set of categories associated with $\mathcal{C}$ only depends by the goals of $\mathcal{C}$

[2]The helpfulness is a measures of the utility of the rating of a review is useful in making a decision and it is computed as the average of the scores.

A. The agents' knowledge representation

Interests and preferences of each owner (i.e., u, g) are taken into account by the associated agent $a \in \mathcal{A}$ and stored into its *profile* p_a consisting of (*i*) interests, (*ii*) behaviors, (*iii*) access modes and (*iv*) trust levels, as follows:

Interests: The interest of the agent a in a category x is a function $I_a(x) : \mathcal{A} \times \mathcal{X} \rightarrow [0, 1] \in \mathbb{R}$ computed as the overall ratio of reviews for items belonging to x. More formally, $I_a(x)$ is expressed by $I_a(x) = |\{r_{a,i} : i \in x\}|/|\overline{\mathbf{r}}_a|$, where $\overline{\mathbf{r}}_a$ is the review history of a, and we denote by $r_{a,i}$ each review contained in $\overline{\mathbf{r}}_a$ and referred to an item i.

Behaviors: The behavior field informs us about the activities admitted or not within a group. It is assumed to be a statement of the form "The average rating of items is greater than 3.0" and so on. Targets (e.g., average rating, frequency of posts, etc.) and goals (e.g., discriminating thresholds) of the statements depend on the specific considered community. Therefore, let b be a behavior (e.g., performed by a user, admitted into a group) and let $\mathcal{B} = \{b_1, b_2, \ldots, b_p\}$ be a given a set of behaviors. We assume to dispose of a function $\zeta_a(b) : \mathcal{A} \times \mathcal{B} \rightarrow \{\texttt{True}, \texttt{False}\}$ which takes an agent $a \in \mathcal{A}$ and a behavior $b \in \mathcal{B}$ and checks whether the behavior b matches with the a's past behaviors. The set of behaviors associated with an agent a will be defined as B_a, i.e., we set $B_a = \{\zeta_a(b) | \quad b \in \mathcal{B}\}$.

Access modes: An access mode identifies a modality for *accessing/allowing the access* to a group, arbitrarily set by the agent owner, like to *open*, *closed* or *secret*, and let L be a list specifying such accessing modes. More in general, we supposed that a function $M : \mathcal{A} \rightarrow L$ is available to associate an agent $a \in \mathcal{A}$ with a mode $l \in L$ for accessing to a group. This components of the user profiles can be considered as fully random, and not correlated with the other components.

Trust levels: We suppose that an *asymmetric trust function* returning how much an agent j perceives another agent k as trustworthy is available (i.e., in general if j trusts k it does not mean that k trusts h too, $t_{j \rightarrow k} \neq t_{k \rightarrow j}$). Trust levels are assigned during the agent interactions and mostly depend on the specific community. For instance, in CIAO everyone can explicitly declare the own trust about each other member. Similarly, the trust perceived by an agent a_u with respect to a group g of agents can be defined as $t_{a_u \rightarrow g} = \sum_{v \in g} t_{a_u \rightarrow a_v}/|g|$.

Note that for a group g: (*i*) its administrator sets the admitted behaviors access modalities (denoted as M_g); (*ii*) the interest for a category $x \in X$ is computed as the average of the interests of its agent members for x that are stored in their profiles; (*iii*) how the members of g perceive as trustworthy an agent a_u is computed as $t_{g \rightarrow a_u} = \sum_{a_v \in g} t_{a_v \rightarrow a_u}/|g|$.

B. The agents' tasks

An agent updates its profile when an action involving information stored therein is performed. More precisely:

- After each performed action, an agent a_u updates in its profile the interests and the boolean values of the involved behavioral variables. Similarly, an agent a_g in its profile updates the behavioral variables every time the administrator of g changes the associated rules. Besides, each time that the preferred access modes change then the associated agent updates its profile.
- When u expresses his/her evaluation about another user v then a_u updates the trust measure.

We also assume that a *Distributed Directory Facilitator* agent (DDF) supports the other agents in $\mathcal{C}$ with an *Agent Indexing Service* and a *Communication Layer* enabling the agent message exchange.

IV. Similarity and Trust

To investigate the time-stability of groups in terms of similarity, by means of the model previously presented, we consider agents' similarities and trust relationships, as follows.

A. Similarity measure

Let $s_{u,v}$ be the measure of similarity between the profiles of agents a_u and a_v computed as a weighted mean of the contributions of interests (c_I), behaviors (c_B) and access modes (c_M) normalized in $[0, 1]$. More formally,

$$s_{a_u,a_v} = (w_I \cdot c_I + w_B \cdot c_B + w_M \cdot c_M)/(w_I + w_B + w_M)$$

where $w_I, w_B, w_M \in [0, 1] \in \mathbb{R}$ are system weights for the contributes c_I, c_B and c_M, in turn, are computed as follows:

- c_I is based on the average difference between the interest values of a_u and a_v for each category $x \in \mathcal{X}$:
$$c_I = 1 - \sum_{c \in C} |I_{a_u}(x) - I_{a_v}(x)|/|\mathcal{X}|$$
- c_B is computed on the average difference between the boolean variables contained in B_{a_u} and B_{a_v}. This difference is $0/1$ if the two corresponding variables are equal/different.
- c_A is set to $1/0$ if M_{a_u} is equal/different to M_{a_u}.

The similarity $s_{u,g}$ between an agent and a group is computed in the same manner described above, simply by substituting a_v with a_g.

B. Trust and compactness measures

We view trust as two terms specifying how much an agent trusts another agent, and how much the community perceives an agent as trustworthy. A *feedback mechanism* usually allows each agent to record its satisfaction for its interactions with other agents in order to compute/update its trust measures [24]–[26] based on the concept of social capital [27]. In fact, a high rate of positive interactions means that an agent can receive an advantage to interact with another agent and, therefore, trust should increase/decrease in presence of positive/negative interactions. The first component, known in the *trust theory* as *reliability*, represents the satisfaction of a_u about a_v, i.e. $rel_{a_u \to a_v}$, and can specify several types of trust relationships (e.g., the *honesty*, the *dependability* or, as in this work, how much a_u is satisfied by the services provided by a_v). Usually, reliability can assume values ranging in $[0..1] \in \mathbb{R}$ and the higher $rel_{a_u \to a_v}$, the higher the perception of the reliability of a_v by a_u. Note that reliability is an asymmetric measure. The second trust component, named *reputation* and denoted it by rep_a in the interval $[0..1] \in \mathbb{R}$, is a global measure of the trust perceived by the whole community about each other agent. The reputation is computed by averaging all the reliability values $rel_{a_u \to a_v}$ for each $a_v \in \mathcal{A}$.

The two trust components are joined in a unique value to compute the *trust* a_u about a_v as $t_{a_u \to a_v} = \alpha_{a_u} \cdot rel_{a_u \to a_v} + (1-\alpha_{a_u}) \cdot rep_{a_v}$, where $\alpha_{a_u} \in [0..1] \in \mathbb{R}$ is set by a_u to weight the relevance it assigns to the first trust term with respect to the second one. Note that trust is an asymmetric measure because in the formulation it takes into account the reliability. Besides, each time a reliability value is updated by a_u, it sends the new value to the DF that, in turn, returns a reputation value to a_u when it needs to compute a trust measure.

As defined in [17], compactness is a measure combining trust and similarity, say $\gamma_{a_u \to a_v}$, able to exploit importance given to the mutual similarity with respect to the mutual trust. We model this level of importance by the coefficient Ws, ranging in $[0..1] \in \mathbb{R}$ and, consequently, we define $\gamma_{a_u \to a_v}$ as $\gamma_{a_u \to a_v} = Ws \cdot s_{a_u,a_v} + (1 - Ws) \cdot t_{a_u \to a_v}$. Remember that trust is an asymmetric measure $\gamma_{u \to v} \neq \gamma_{v \to u}$.

In Table I the meaning of the symbols is reported.

TABLE I
Main symbols used in the paper and their meaning.

Symbol	Meaning
$t_{u \to v}$	level of trust perceived by the user u w.r.t. the user v
$t_{u \to g}$	level of trust perceived by user u w.r.t. the group g
$t_{g \to u}$	level of trust perceived by the group g w.r.t. the user u
$s_{u,v}$	similarity between users u and v
$s_{u,g}$	similarity between the user u and the group g
$rel_{u \to v}$	reliability perceived by the user u w.r.t. a_v
rep_u	reputation of the user u
α_u	weight that a_u assigns to the reliability w.r.t. the reputation
$\gamma_{u \to v}$	compactness perceived by a_u w.r.t. a_v
Ws_u	weight that a_u assigns to the similarity w.r.t. the trust
W_I	weights the interest in computing $s_{u,v}$
W_B	weights the behavior in computing $s_{u,v}$
W_M	weights the access mode in computing $s_{u,v}$

V. A2G: Matching Agents with Groups

In this section the algorithm (*A2G*), enabling user agents to select the groups to join with, is presented.

Let $\mathcal{G} = \{g_1, g_2, \ldots, g_n\}$ be the groups of $\mathcal{C}$, with $|\mathcal{G}| = n$. Moreover, let $k^{a_u}_{\mathtt{MAX}}$ be the upper bound ranging in $[0, n]$ which specifies the number of groups a_u desires to join with and reasonably it will be $k_{\mathtt{MAX}^{a_u}} << n$. In the following, for convenience, the notation $k_{\mathtt{MAX}}$ will be used instead of $k^{a_u}_{\mathtt{MAX}}$.

Algorithm A2G selects $k_{\mathtt{MAX}}$ groups having the largest value of compactness of a_u vs the joined groups. We assume that a_u continues to receive the whole benefit from all the $\mathcal{K} \subseteq \mathcal{G}$ groups which it is joined with, so that the overall received benefit in joining with all the $\mathcal{K}$ groups, in term of compactness, is given by $\sum_{g_i \in \mathcal{K}} \gamma_{u \to g_i}$. Finding the subset $\mathcal{K}^\star \subseteq \mathcal{G}$ producing the best benefit for a_u under the constraint

$|\mathcal{K}^\star| = k_{\mathtt{MAX}}$ is equivalent to solve an optimization problem. In this work, we assume that each user agent a_u is unable to know, in advance, the compactness of all groups in $\mathcal{G}$. Furthermore, we assume that: (*i*) a_u is able to sample m random groups from $\mathcal{G}$; (*ii*) a_u will record into an internal cache, denoted as H the profiles of the groups a_u joined with in the past; (*iii*) m is the number of the group agents that at each epoch must be contacted by a_u. Algorithm 1 describes the steps a_u performs to find the $k_{\mathtt{MAX}}$ groups it can join with, while the Algorithm 2 runs on the group agent. In particular, it is assumed that (*i*) the size of each group $g \in \mathcal{G}$ is $\leq$ than a threshold $n_{\mathtt{MAX}}$; (*ii*) $n_{\mathtt{MAX}}$ is fixed by the group administrator; (*iii*) each agent a_g stores into an internal cache the profiles of the agents who joined with g;

VI. The proposed computational framework

In our framework, a community is associated with a temporal dataset of events consisting of a matrix EM, where each row represents an event containing a timestamp, an agent identifiers and the event attributes. An events can be an action performed by an agent or an external event changing the state of the community. Furthermore, we assume that a (non time-varying) matrix TM of trust relationships is available, where each row is a pair of agent IDs (a_u, a_v) representing a trust relationship among agent a_u and agent a_v. Moreover, *A2G-Comp* and *A2G-Sim* are two versions of the algorithm A2G. In the former the compactness γ is computed by setting $Ws < 1$ (i.e., it is driven by similarity and trust) and for the latter $Ws = 1$ (i.e., it is driven only by similarity).

The framework provides the weights w_I and Ws in the range $[0, 1] \in \mathbb{R}$ influencing the A2G algorithm results. w_I represents the weight assigned to the agent interest, while $1 - w_I$ is divided between w_B, representing the weight assigned to the agents behavior (see Section IV), and w_M. c_M is set as *random* not being into the original dataset. The lower the value of w_I, the higher the incidence of the component c_M in computing the similarity, in other words c_M is an “uncorrelated component”. Furthermore, the higher the value of Ws, the lower the impact of the trust relationship in computing the compactness γ.

In this perspective, the MAC (Mean Average Compactness) and MAS (Mean Average Similarity) measures have been used to perform experiments, in dependence of w_I and Ws.

Data: a_u: an agent, H the current set of groups of a_u, m: an integer in $[0, n]$, $k_{\mathtt{MAX}}$: the number of groups a_u can join with
Result: The new set S of groups of a_u
Y = a set of m random groups extracted from DF; $Z = H \bigcup Y$;
for $(g \in Z)$ $\{a_u$ sends a message to a_g and let p_g be the profile of $g\}$;
$\mathcal{S}$ = the set of $k_{\mathtt{MAX}}$ groups of Z with the highest compactness values;
for $(g \in \mathcal{S})$ **do**
  if $(g \notin H)$ **then** $\{a_u$ sends a join request with its profile to $a_g\}$;
  else $\{a_u$ left $g\}$;
end
return S

Algorithm 1: The User Agent Task

Data: r: an agent which has asked to join with the group, K, the set of agents in g
for $(a_u \in K)$ **do** $\{a_g$ sends a message to $a_u\}$;
for $(a_u \in K \bigcup \{r\})$ **do** $\{$Compute $\gamma_{g \to u}\}$;
Let $\pi = \dfrac{\sum_{u_i \in g} \sum_{u_j \in g} \gamma_{u_i \to u_j}}{|g|^2}$ $\quad \forall \langle u_i, u_j \rangle \in g$ and let $\mathcal{S} = \varnothing$;
for $(u \in K \bigcup \{r\})$ **do** $\{$**if** $(\gamma_{g \to u} \geq \pi)$ $\{\mathcal{S} = \mathcal{S} \bigcup \{u\}\}\}$;
Let $\mathtt{Top}_{\mathcal{S}}$ be the set of top-$n_{\mathtt{MAX}}$ users in $\mathcal{S}$;
if $(r \in \mathcal{S})$ $\{a_g$ accepts the join request of $r\}$;
for $(u \in K \wedge u \notin \mathcal{S})$ **do** $\{a_g$ deletes u from $g\}$;

Algorithm 2: The Group Agent Task

$$MAC(w_I, Ws) = \frac{\sum_{g \in G} AC_g}{|G|} \qquad AC_g = \frac{\sum_{x,y \in g, x \neq y} \gamma_{x \to y}}{|g|}. \tag{1}$$

$$MAS(w_I, Ws) = \frac{\sum_{g \in G} AS_g}{|G|} \qquad AS_g = \frac{\sum_{x,y \in g, x \neq y} s_{x \to y}}{|g|}. \tag{2}$$

More in detail, AC_g (resp., AS_g) is defined as the *Average Compactness* (resp., *Average Similarity*), similarly to the average dissimilarity commonly exploited in Clustering Analysis [28], since a group g can be viewed as a cluster of agents. Note that MAC is computed in the training phase of the A2G-Comp algorithm (i.e. it only drives group formation), while MAS is computed in the *test* phase. Therefore measuring the variation of MAS can be useful to verify the homogeneity, in terms of similarity, of the groups formed in the training phase.

A. Experimental approach and main parameters

We considered the time-variation of the average similarity of the groups in two different cases: (*i*) (**Comp**) Groups formed by the *A2G-Comp* algorithm are driven by compactness ($Ws < 1$). (*ii*) (**Sim**) Groups formed by the *A2G-Sim* algorithm are driven by the similarity criterion ($Ws = 1$).

The computation of the measures are performed by following the steps described below: (*i*) Rows of the matrix EM are *arranged in an increasing order*, basing on the database timestamp. (*ii*) The matrix EM is divided into a *number of time-windows* of equal size. The first time-window is for the training set and the remaining for the tests. (*iii*) The *trust network* is built by loading the matrix TM and for all. (*iv*) The *training* is performed by executing the algorithm *A2G-Comp* (resp. *A2G-Sim*) on the first time-window, in order to form groups of agents. (*v*) The training is stopped when “stable” values of MAC for *A2G-Comp* (MAS for *A2G-Sim*) are reach, i.e. the difference between two steps is less than a given threshold (in our case it was 5%). (*vi*) Data of the remaining time-windows is loaded in sequence for each of them and MAS is computed without executing the algorithm A2G, such that group composition remains the same as in the end of the training phase. This technique allows us to study the variation of MAS due to the addition of events representing the execution of some further actions by the agents.

TABLE II
PARAMETERS USED IN EXPERIMENTS ON THE CIAO DATASET.

Parameter	Value	Parameter	Value
Number of Groups	50	$k_{\mathtt{MAX}}$	10
$k_{\mathtt{MIN}}$	0	N_{REQ}	5
$k_{\mathtt{MAX}}$	50	Size of the Training Set	10, 000
$n_{\mathtt{MIN}}$	0	Size of the Test Set	26, 065

VII. EXPERIMENTS

Experiments exploited the described framework on a dataset extracted from the social network CIAO [12][3] referred to $12,375$ users and consisting of two matrices $(i.e., EM, TM)$. Rows of matrix EM have the form $\{$*userID, productID, categoryID, rating*, *helpfulness*, *timestamp*$\}$, where *categoryID* is the commercial category of the item identified by *productID* which received the rating by the reviewer identified by *userID*, and *helpfulness* represents the level of satisfaction of the other member for that *rating* (it has not been used in our experiments). Table VII contains the parameters used to carry out the experiments. The training set is made by the first $10,000$ events. The software used for the experiments can be downloaded at https://github.com/fmes/simU2G.

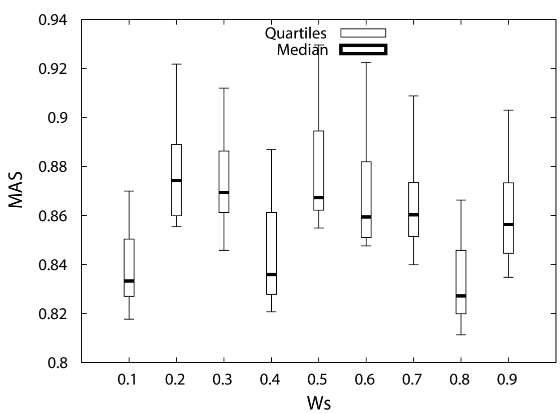

Fig. 1. MAS vs parameter Ws achieved by the A2G-Comp ($w_I = 0.1$).

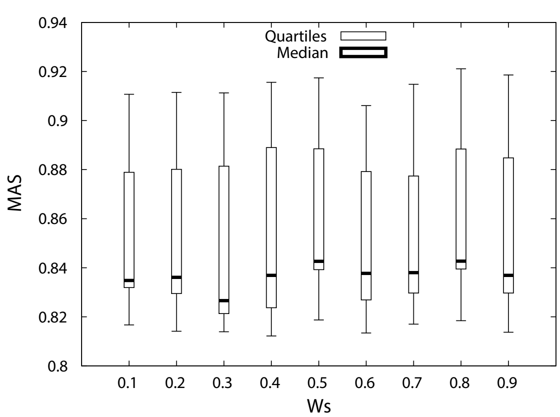

Fig. 2. MAS vs parameter Ws achieved by the A2G-Comp ($w_I = 0.5$).

[3]Data used in our experiments are publicly available at http://www.public.asu.edu/~jtang20/datasetcode/truststudy.htm

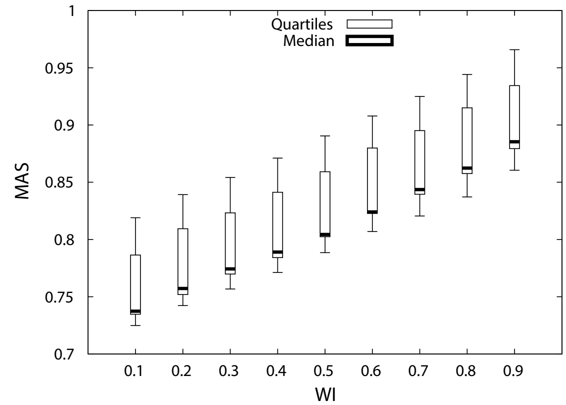

Fig. 3. MAS vs parameter w_I achieved by the A2G-Sim ($W_s = 0.1$).

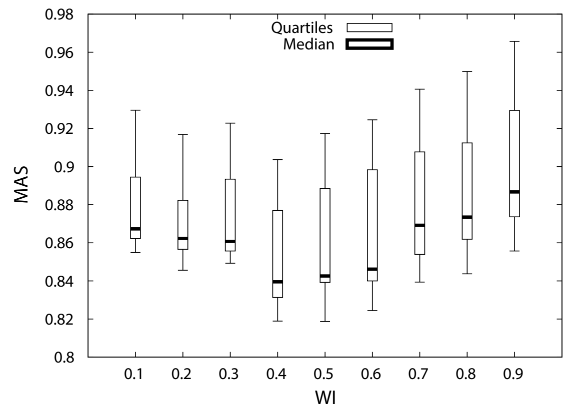

Fig. 4. MAS vs. parameter w_I achieved by the A2G-Comp ($Ws = 0.5$).

A. Results

Experiments were performed by varying weights w_I and Ws in the range $[0.1 - 0.9]$. Figures 1 and 2 report the execution of *A2G-Comp* for $w_I = 0.1$ and $w_I = 0.5$ and Ws in the range $[0.1 - 0.9]$. In Figure 1 the values of the MAS are relevant also when the weight of the trust component is low (i.e., $Ws > 0.5$). In particular, values of MAS are larger than 0.8 (e.g., median is 0.82 for $Ws = 0.8$), giving a good time-stability with a visible bias around the median value, if compared to the results shown in Figure 2, on which the uncorrelated component starts to assume a less significant weight. Figures 3-4 show the results for the 1st, 2nd, and 3rd quartile, minimum and maximum values of MAS computed after the training for the remaining time windows. Figure 3 shows the behavior of *A2G-Sim* for different values of w_I, while Figure 4 represents the different values of MAS for A2G-Comp with $Ws = 0.5$. From these Figures we note that the lower the value of w_I, the lower the value of overall similarity at the end of the test for A2G-Sim (Figure 3), while for *A2G-Comp* the values of MAS are higher of about 10%. This first set of results say us of driving groups formation by compactness when the weight w_M is of at least 25% to obtain homogeneous time-stable groups in terms of average similarity.

VIII. Conclusions

The experimental study has been conducted with a distributed algorithm for groups formation, named A2G, which exploits the compactness measure, i.e. a combination of similarity and trust. The experimental approach of the conceptual framework permits to employ different combination of similarity and trust.

Obtained results have shown that forming groups on the basis of users' similarity will lead to form time-stable homogeneous groups if the weight of the uncorrelated behavioral components is marginal. Nevertheless, when group formation is driven by compactness (i.e., by combining similarity and trust) then groups result time-stable homogeneous even if the uncorrelated components included in the computation of similarity are relevant. Therefore, trust relationships will help to improve the level of resilience, in terms of similarity, also in presence of behavioral components which are not strongly linked with the others. Interestingly, even when the weight assigned to the trust relationship in the computation of compactness is very low, group formation driven by compactness will lead to a number of groups having a higher level of time-stability with respect the similarity measure.

References

[1] R. Kraut and A. Fiore, "The role of founders in building online groups," in *Proc. 17th ACM Conf. CSCW 2014*. ACM Press, 2014, pp. 722–732.

[2] P. Grabowicz, L. Aiello, V. Eguiluz, and A. Jaimes, "Distinguishing topical and social groups based on common identity and bond theory," in *Proc. of the ACM Int. Conf. WSDM 2013*. ACM, 2013, pp. 627–636.

[3] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, "Group Formation in Large Social Networks: Membership, Growth, and Evolution," in *Proc. 12th ACM SIGKDD Int. Conf.* ACM Press, 2006, pp. 44–54.

[4] S. Kairam, D. Wang, and J. Leskovec, "The life and death of online groups: Predicting group growth and longevity," in *Proc. 5th ACM int. conf. WSDM 2012*. ACM Press, 2012, pp. 673–682.

[5] W. Chen, D. Zhang, and E. Chang, "Combinational collaborative filtering for personalized community rrecommendation," in *Proc. of the ACM Int. Conf. SIGKDD'08*. ACM, 2008, pp. 115–123.

[6] J. Gorla, N. Lathia, S. Robertson, and J. Wang, "Probabilistic group recommendation via information iatching," in *Proc. of the Int. Conf. WWW'13*. ACM Press, 2013, pp. 495–504.

[7] V. Vasuki, N. Natarajan, Z. Lu, B. Savas, and I. Dhillon, "Scalable affiliation recommendation using auxiliary networks," *ACM Trans. on Intelligent Systems and Technology*, vol. 3, no. 1, p. 3, 2011.

[8] A. Comi, L. Fotia, F. Messina, G. Pappalardo, D. Rosaci, and G. M. L. Sarné, "Supporting knowledge sharing in heterogeneous social network thematic groups," in *Complex, Intelligent, and Software Intensive Systems (CISIS), 2015 Ninth International Conference on*. IEEE, 2015, pp. 480–485.

[9] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, "Group formation in large social networks: membership, growth, and evolution," in *Proc. 12th ACM Int. Conf. SIGKDD'06*. ACM, 2006, pp. 44–54.

[10] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proc. 7th ACM Conf. SIGCOMM'07*. ACM, 2007, pp. 29–42.

[11] C. Lortie and M. Guitton, "Looking similar promotes group stability in a game-based virtual community," *GAMES FOR HEALTH: Research, Development, and Clinical Appl.*, vol. 1, no. 4, pp. 274–278, 2012.

[12] J. Tang, X. Hu, H. Gao, and H. Liu, "Exploiting local and global social context for recommendation," in *Proc. of the Int. Joint Conf. IJCAI'13*. AAAI Press, 2013, pp. 2712–2718.

[13] D. Rosaci, "Trust measures for competitive agents," *Knowledge-based Systems (KBS)*, vol. 28, no. 46, pp. 38–46, 2012.

[14] D. Rosaci and G. M. L. Sarné, "Matching users with groups in social networks," in *Intelligent Distributed Computing VII*. Springer, 2014, pp. 45–54.

[15] A. Comi, L. Fotia, F. Messina, D. Rosaci, and G. M. Sarné, "Grouptrust: Finding trust-based group structures in social communities," in *International Symposium on Intelligent and Distributed Computing*. Springer, 2016, pp. 143–152.

[16] W. Sherchan, S. Nepal, and C. Paris, "A survey of trust in social networks," *ACM Computing Surveys*, vol. 45, no. 4, p. 47, 2013.

[17] P. De Meo, E. Ferrara, D. Rosaci, and G. M. L. Sarné, "Trust and compactness in social network groups," *Cybernetics, IEEE Transactions on*, vol. 45, no. 2, Feb 2015.

[18] A. Comi, L. Fotia, F. Messina, G. Pappalardo, D. Rosaci, and G. M. L. Sarné, "Forming homogeneous classes for e-learning in a social network scenario," in *Intelligent Distributed Computing IX*. Springer, 2016, pp. 131–141.

[19] P. De Meo, E. Ferrara, D. Rosaci, and G. M. L. Sarné, "How to improve group homogeneity in online social networks," in *Proceedings of the 14th Workshop from "Objects to Agents", WOA 2013*, ser. CEUR Workshop Proceedings, vol. 1099. CEUR-WS, 2013, pp. 73–77.

[20] P. De Meo, F. Messina, D. Rosaci, and G. M. L. Sarné, "Improving the compactness in social network thematic groups by exploiting a multi-dimensional user-to-group matching algorithm," in *Intelligent Networking and Collaborative Systems (INCoS), 2014 International Conference on*. IEEE, 2014, pp. 57–64.

[21] D. Prentice, D. Miller, and J. Lightdale, "Asymmetries in attachments to groups and to their members: Distinguishing between common-identity and common-bond groups," *Personality and Social Psychology Bulletin*, vol. 20, no. 5, pp. 484–493, 1994.

[22] H. Purohit, Y. Ruan, D. Fuhry, S. Parthasarathy, and A. Sheth, "On the role of social identity and cohesion in characterizing online social communities," *arXiv preprint arXiv:1212.0141*, 2012.

[23] J. Nahapiet and S. Ghoshal, "Social capital, intellectual capital, and the organizational advantage," *Academy management review*, vol. 23, no. 2, pp. 242–266, 1998.

[24] T. DuBois, J. Golbeck, and A. Srinivasan, "Predicting trust and distrust in social networks," in *PASSAT'11 IEEE 3rd Int. Conf. on and SocialCom'11 IEEE 3rd Int. Conf.* IEEE, 2011, pp. 418–424.

[25] H. Liu, E. Lim, H. W. Lauw, M. Le, A. Sun, J. Srivastava, and Y. Kim, "Predicting trusts among users of online communities: an epinions case study," in *Proc. 9th ACM conf. on E. com.* ACM, 2008, pp. 310–319.

[26] P. De Meo, F. Messina, D. Rosaci, and G. M. L. Sarné, "Recommending users in social networks by integrating local and global reputation," in *International Conference on Internet and Distributed Computing Systems*. Springer, 2014, pp. 437–446.

[27] P. S. Adler and S.-W. Kwon, "Social capital: Prospects for a new concept," *Academy management review*, vol. 27, no. 1, pp. 17–40, 2002.

[28] R. K. Pearson, T. Zylkin, J. S. Schwaber, and G. E. Gonye, "Quantitative evaluation of clustering results using computational negative controls," in *Proc. of the 2004 SIAM International Conference on Data Mining*, 2004, pp. 188–199.