Motivating Students in Collaborative Activities With Game-Theoretic Group Recommendations

Zacharoula Papamitsiou, *Member, IEEE* and Anastasios A. Economides

***Abstract*—Recommending educational resources to groups of students is a common task in collaborative learning contexts. However, differences in within-group motivational factors might lead to conflicts in students' intention to use the resources. Previous methods fail to achieve high goodness of recommendation for the majority of students in heterogeneous groups. This study demonstrates a game-theoretic solution for recommending educational resources to homogeneous and heterogeneous groups. The group members are the players, the resources comprise the set of possible actions, and selecting those items that will maximize all students' motivation in the collaborative activity is a problem of finding the Nash Equilibrium (NE). In case the NE is Pareto efficient, none of the players can get more payoff (motivation) without decreasing the payoff of any other player, indicating an optimal benefit for the group as a whole. The suggested approach was empirically evaluated in a controlled experiment with a real dataset. The relevance of each delivered item to its corresponding students was explored both from the perspective of the group and its the individual students. The accuracy of the predicted group/individual motivation, the goodness of the ranked list of recommendations, and the problem-solving performance for the treatment group were significantly higher compared to the control groups. Limitations of the approach, as well as future work plans conclude the paper.**

***Index Terms*—Collaborative learning, game-theory, group recommendation, motivation, noncooperative games.**

I. Introduction

RECOMMENDING educational resources to groups of students instead of individuals is a common practice in collaborative learning contexts. However, delivering those resources that will prompt all students' thinking and learning, motivate all group members to actively engage in the collaborative activity, and accomplish their learning needs is a nontrivial task [1], [2]. The reason behind this claim is that students in a group may not be fulfilled by the same items, yet wish to meet their own learning goals, making it difficult for the group to reach a consensus regarding the consistency between the expected gain from the resources and the actual experience, as well as their intention to use the resources. When students participate in groups, each student is influenced by the perceptions, decisions, and choices of the other students, and they all together try to agree upon what is considered as useful, helpful, and whether it facilitates the activity's requirements and their own learning objectives.

Manuscript received January 25, 2018; revised August 11, 2018; accepted September 3, 2018. Date of publication September 9, 2018; date of current version June 18, 2020. *(Corresponding author: Zacharoula Papamitsiou.)*

The authors are with the University of Macedonia, Information Systems Program, Thessaloniki 546 36, Greece (e-mail: papamits@ uom.edu.gr; economid@uom.gr).

Digital Object Identifier 10.1109/TLT.2018.2869582

The problem this study addresses is how to optimally recommend educational resources to groups of students with respect to each individual member's motivation and intention to use the recommendation.

Inspired from [3], we argue that *Game Theory* could efficiently solve "conflicts of interest" between the group members [4] and guide the recommendation of educational resources. In the collaborative learning context, "conflict of interest" occurs when individual perceptions about the value of educational resources on self-determined learning goals (i.e., personal benefits) are potentially competing to the benefits of the other group members. It implies the diversity in group members' self-motivated considerations regarding the usefulness of the resources, as well as in their behavioral intention to finally use the resources.

Game theory is "a study of mathematical models of conflict and cooperation between intelligent rational decision-makers (players)" [5]. This study demonstrates a method for recommending educational resources to groups of students based on noncooperative games. *Noncooperative* is a technical term and not an assessment of the degree of cooperation among players in the game; a noncooperative game can model cooperation, focusing on predicting individual players' actions and payoffs, but the players make self-enforced decisions independently [4].

The remainder of the paper is organized as follows. Section II briefly reviews existing methods for group recommendations, highlighting the need for additional research in the educational domain. Section III formulates the problem of recommendation of educational resources to groups of students as a noncooperative game. Section IV explains the experimental methodology for the evaluation of the suggested approach, and Section V demonstrates the empirical results. Section VI elaborates on the findings and Section VII outlines the contributions and limitations, and concludes the paper.

II. Related Work, Motivation of the Research, and Research Question

Recommender systems (RSs) are programs that search in collections of items (e.g., products, services or people) and target at suggesting to their users those items that will best satisfy their preferences, by inferring the users' potential interest in the items [6]; the decision is driven by the collected and analyzed information about the items, the users and the user-item interactions [7].

The level of relevance of an item to a user is usually expressed by the degree of user's appreciation on it, i.e., a rating. Literature on the topic is rich [7]–[9]. Prevalent approaches include information filtering systems [63], collaborative filtering [10], content-based [11], and knowledge-based [12] techniques. Due to drawbacks and limitations of these techniques (e.g., prediction accuracy, data sparsity, and cold-start issues), more sophisticated approaches have been proposed, e.g., fuzzy-logic based [13], social network-based [14], and context-aware RSs [15].

A. Group RSs

In many cases, in different application domains, the users have to carry out an activity together, as a group, resulting in a rapid increase of RSs that cope with the challenge of addressing recommendations for groups of users instead of individuals. In these cases, the goal is to recommend items that would meet all users' preferences as much as possible [15], [58]. However, recommending to groups is more complex than recommending to individuals [2]; group members usually do not have the same preferences and interests, making it difficult to reach an agreement between them and satisfy them all. Dissimilarities among users are apparent in group recommenders [16], and as a result, how the group members' disagreements on the same items are bridged is critical for the effectiveness of group recommendations [17].

In order to address issues of disagreement between the group members, the researchers from the group RSs domain adopted the semantics originating from theories that consider social groups, such as Planned Behavior Theory [61], Social Learning Theory [62], and the Social Choice Theory [18], [19]. Methods such as popularity voting, most respected person, least misery (LM), most pleasure (MP), and average [2] have been applied to the decision-making process to solve conflicting ratings of preferences and to establish an automatic way of how a group of people can reach to consensus [20]. For example, MusicFX, a group RS for selecting a music station, uses a variant of the average without misery strategy for group profile aggregation [21]. INTRIGUE, a hybrid system for sightseeing destination recommendation to tourists, takes into account subgroups' characteristics and aggregates the subgroup recommendations according to the created models [22]. PolyLens recommends movies to groups by fusing recommendations using the LM criterion [23], whereas HappyMovie uses the individuals' "social trust" and personality in an average profile strategy [24]. Other works focus on integrating the social, expertise, and preference dissimilarity in the recommendation process [25]. For a systematic review, see [26].

In the educational group RSs research field, and, although still rather sparse, the prevalent approaches include: 1) merging individual preferences in a pseudo group profile according to an aggregation strategy, prior to generating the recommendation (e.g., [27], [28]), 2) constructing groups with high inner member similarity (homogeneous) and recommending resources from a merged list of recommendations, generated for each group member individually (e.g., [29], [30]), or 3) evaluating aggregation methods and applying classification on meta-data, including the prior evaluation results and a set of learners' characteristics [31]. In these approaches, the consensus functions aggregate the group members' personalities, interests, and learning styles.

B. Motivation of the Research—Research Questions

However, in these methods, four common types of drawbacks are related to: 1) the group inner similarity, 2) the aggregation strategies, 3) the number of recommended resources, and 4) the individual members' conformity degree, i.e., the adjustment of one's opinion toward the majority seeking for approval [32]. Regarding the first drawback type, homogeneous groups is an unwanted restriction, since homogeneity in group synthesis is not always possible to achieve. Moreover, heterogeneity in groups is considered as more beneficial for learners in collaborative learning contexts [33], [34]. In addition, studies in cognitive psychology state that in processes that concern judgments (such as decision making in a group recommendation process), both concepts of similarity and dissimilarity share equal importance [35], [36]. Furthermore, the existing methods fail to achieve high-quality performance and goodness of recommendation for the majority of students in heterogeneous groups [26], [37]. Regarding the second drawback type, not all aggregation strategies work efficiently in all cases [31], whereas evaluating the aggregation strategies prior to applying one of them is time consuming (if not raising a fairness issue in recommendation). Regarding the third drawback type, existing methods recommend only one item at a time, although it is very likely that students want to access multiple learning resources. In this case, they would be more pleased with a sequence of suggested items. Finally, regarding the fourth drawback type, the focus of recommendation is on the overall group satisfaction, bypassing the relevance of the recommended items to the corresponding students, and how beneficial these items are to the learning subjects themselves. To the best of our knowledge, none of the abovementioned approaches consider in the recommendation process the students' behavioral intention to take the recommendation and use the resources. Similarly, measuring students' persistence—indicating the actual engagement with the resources (in a learning analytics fashion)—is missing.

Thus, the emerging research questions are as follows.

RQ1: Can we accurately and efficiently recommend sequences of educational resources to homogeneous and heterogeneous groups of students, with respect to both the individuals' and the group's motivation, and intention to use the resources?

RQ2: What is the impact of a recommendation on individual students' persistence as well as on the groups' learning performance in the collaborative problem-solving activity?

Toward answering the research questions, we argue that noncooperative games could efficiently solve conflicts of interest between group members and guide the recommendation of a sequence of learning resources.

III. Problem Formulation as Noncooperative Game

We examine the case of having students who collaboratively solve problems in groups (at least two members). The group members are students with *potentially conflicting* self-determined learning goals and expectations (i.e., motivation to be engaged in the collaborative problem-based activity and perceive as useful the suggested items). Thus, the groups might be homogeneous (high similarity between the group members), mildly heterogeneous (medium similarity between the group members), or heterogeneous groups (low similarity between the group members). The goal is to recommend to each group those items (single or sequence) that will be beneficial to the group as a whole—supporting the group members to efficiently complete the assigned collaborative activity—and that are expected to maximize all members' motivation to take the recommendation and use the items, and their actual engagement with the items, i.e., persistence.

A. Problem Definition

Consider a set of students $L = \{l_i\}$ and a set of educational resources $R = \{r_j\}$; the indexes i, j refer to an individual student or resource (item), respectively. Let G be a set of all groups that may be formed by L; then,$|G| = 2^n - n - 1$. If $g \in G$, then, $|g| = k$, the number k of group members in group g, with $k \geq 2$. For each student i and each item j, the student's motivation to take the recommendation m_{ij} can be estimated from the student's self-enforced evaluation of how much the particular item corresponds to the student's intrinsic interest and satisfaction, and how useful the student considers the item to be (intrinsic motivation), with respect to the student's goals and achievement expectations (extrinsic motivation). Student's motivation triggered by each item is measured with an appropriate questionnaire in a 5-point Likert-like scale (briefly outlined in Section III-C); if a student has not yet evaluated an item, then $m_{ij} = 0$.

Also, consider $\hat{m}_{ij}$ as the predicted motivational excitement that item j imparts to student i, i.e., a *decision criterion* for selecting the item, approximated with the Matrix Factorization technique [38] (briefly explained in Section III-D).

The items to be recommended to each group should not have been previously seen or evaluated by any of the group members, and are strategically designed to promote the students' interest in a specific learning topic and improve their competences accordingly.

We model the group recommendation problem as a noncooperative game, i.e., a triad $(k,\ Q,\ f)$ where

1) the k students (group members) are the *players*;
2) the set of unrated items $Q = \{x_z\} = \bigcap_i \{j|\hat{m}_{ij} = 0\}$, $Q \subseteq R$, are the available *actions*; a vector $x = (x_1, \ldots, x_\mu) \in Q$ is a *strategy profile*, which includes the actions of all players;
3) the *payoff* function for a student i and a strategy profile x calculates the predicted motivation of student i in the group, resulting from the actions by all group members—including himself—as the average individual predicted motivation from all items in x; $payoff$ is computed by

$$f_i(x) = \frac{\sum_z \hat{m}_{iz}}{|q|}$$

Algorithm 1: Finding the NE and selecting the best-response strategy

Input: k, Q, f (The non-cooperative game)
Output: $x* \in Q$ (The NE profile strategy)

for i = 1; i $\leq$ k; i + + **do** //for all students
  assign $x_i \in Q, x_{-i} \in Q$ //initialize profile strategy
repeat
  repeat
    for i = 1; i $\leq$ k; i + +**do** //for all students
      assign $x_i* \in Q, x_{-i}* \in Q$ //assign another strategy
        compute $f_i(x_i^*, x_{-i}^*)$ //compute the payoff
  until $f_i(x_i^*, x_{-i}^*) \geq f_i(x_i, x_{-i}^*)$ //no student has incentive //to change the strategy
  if $f_i(x*) > f_i(x)$ **then** //check for Pareto efficiency
      Pareto = true
  compute d_i // difference max-min in NE
until Pareto or min d_i
return $x* \in Q$ //the NE

where $|q|$ is the total number of items in x.

The items that will be recommended to the group of students are those in the nash equilibrium (NE) (single item or sequence of items). A strategy profile $x* \in Q$ is an NE if

$$\forall i, x_i \in Q_i, x_i \neq x_i^* : f_i(x_i*, x*_{-i}) \geq f_i(x_i, x*_{-i})$$

where x_i is a strategy profile for student i and x_{-i} is a strategy profile of all students except for student i. In other words, considering that the other students will not modify their own strategy, the student who has the option of deviating should have no benefit by unilaterally changing his own strategy. In the group recommendation problem, in the NE, no student i can further increase their motivation from the recommendation by altering their strategy to $x_i \neq x_i^*$, provided that all other students stay with their selected strategies. The students' strategies converge to the NE after an iterative best-response strategy update.

In case there are more than one strategies that are NE, the recommendation solution for this group is the one that is "*socially optimum*": no other strategy $x* \in Q$ has both a weakly better payoff for all students and a strictly better payoff for some student: $\forall i\ f_i(x*) \geq f_i(x)$ and $\exists\, \mathrm{if}_i(x*) > f_i(x)$. In other words, if the recommendation solution is Pareto optimal, it is impossible to improve the motivation of a student without worsening the motivation of another student, indicating an optimal solution for the group as a whole.

However, it is very possible that none of the NE is Pareto optimal. In this case, we calculate the distance between the highest and lowest payoffs in the strategies that are NE, and select the strategy that minimizes this distance, indicating a fair solution for the group.

Algorithm 1 presents the algorithm for finding the NE and selecting the best-response strategy.

Finally, for calculating the predicted group motivation and intention to use the recommendation, a group consensus function $M(g, x)$ computes the average motivation from each item in the recommended strategy x for the group g, where $f_i(x)$ is the payoff for each member i

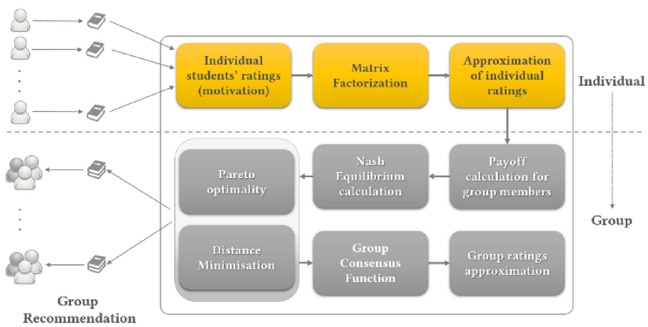

Fig. 1. Architecture of the noncooperative GT group recommender system for educational resources.

TABLE I
INDIVIDUAL STUDENTS' PREDICTED MOTIVATION FROM THE EDUCATIONAL RESOURCES

	r1	r2	r3	r4	r5
A	3.6	4.2	1.8	2.6	3.2
B	1.2	3.4	2.4	4.6	4.6

$$M(g,x) = \frac{\sum_{x\in Q} f_i(x)}{|g|}.$$

The architecture of the suggested approach for educational group recommendations is illustrated in Fig. 1.

B. Illustrative Example

Consider a group with two members, i.e., the students A and B. Table I demonstrates the individual students' predicted motivation $\hat{m}_{ij}$ to be excited from the educational resources r1, r2, r3, r4, and r5, after matrix factorization. None of the students A and B has previously seen or evaluated any of these five items.

Table II illustrates the payoff (motivation) for each student from all the possible actions (strategies) taken by himself and the other group member.

From this table, it can be seen that there are two NE, the strategies profiles (r2, r4) and (r2, r5). The reason is that if student A chooses action r2, then student B has the same benefit from actions r4 and r5, and does not benefit in changing his action to r1 or r2 or r3. Likewise, considering that student B chooses action r4 or r5, then student A has no benefit to change the action from r2 to r1 or r3 or r4 or r5. Between these two strategies, (r2, r5) is Pareto optimal. This means that these two items (r2 and r5) should be recommended to the group members in order to optimize the motivation for each individual member, whereas, neither of the students can get more payoff (motivation) without decreasing the payoff of the other student, indicating an optimal solution for the group as a whole.

C. Measuring Motivation Excited From Resources

Motivation is defined as "the process whereby goal-directed activity is instigated and sustained" [39] (p. 5). In the context of learning, motivation is "a student's tendency to find academic activities meaningful and worthwhile and to try to derive academic benefits from them" [40] (p. 249). Many theories have been proposed to measure different motivational constructs (e.g., expectancy value theory [41], attribution theory of achievement motivation [42], goal-orientation theory [43], and self-determination theory [44]) and explain why it critically affects learning.

TABLE II
THE PAYOFF (MOTIVATION) FOR EACH STUDENT FROM ALL THE POSSIBLE ACTIONS (STRATEGIES)

		B				
		r1	**r2**	**r3**	**r4**	**r5**
A	**r1**	(3.6, 1.2)	(3.9, 2.3)	(2.7, 1.8)	(3.1, 2.9)	(3.4, 2.9)
	r2	(3.9, 2.3)	(4.2, 3.4)	(3.0, 2.9)	(3.4, 4.0)	(3.7, 4.0)
	r3	(2.7, 1.8)	(3.0, 2.9)	(1.8, 2.4)	(2.2, 3.5)	(2.5, 3.5)
	r4	(3.1, 2.9)	(3.4, 4.0)	(2.2, 3.5)	(2.6, 4.6)	(2.9, 4.6)
	r5	(3.4, 2.9)	(3.7, 4.0)	(2.5, 3.5)	(2.9, 4.6)	(3.2, 4.6)

TABLE III
QUESTIONS FOR MEASURING MOTIVATION

Question
Clarity of item's content: *The item was clear and understandable*
Goal fullfilment from the item: *The item met my expectations and covered my learning goals*
Usefulness of the item: *The item helped me improve my learning*
Behavioral intention to use the item: *I indent to use the item in the future*

The definition of motivation highlights that motivation is a goal-oriented process [39]; it determines the students' goal-setting process, affecting their choices and decisions, accordingly [50]. Moreover, previous research results show that when a system triggers the students' intrinsic motivation (i.e., is perceived as challenging and satisfying [46]), then, the students' behavioral intention to use that system increases (e.g., [47]). In addition, Davis [48] showed that perceived usefulness is an example of extrinsic motivation (i.e., students are doing an activity for its instrumental value [46]) for intention to use information services. Furthermore, it was found that the clarity of educational content could affect learner perceptions of usefulness [49], and as such, clarity could be indirectly considered as a factor that motivates the e-learner.

According to the above, and to keep the measurement of motivation simple, yet contextualized and coherent to the research, for assessing their appreciation to each item, the students had to rate 1) their own perceived clarity of each item, 2) how much each item fullfield their own learning goals, 3) their own perceived usefulness of each item, and 4) their intention to use the item. For this purpose, four questions were delivered to them in a 5-point Likert-like scale (see Table III). The average score per student was considered as the student's perceived motivation from the corresponding item.

D. Motivation Approximation: Matrix Factorization

By definition, in noncooperative games, the students act rationally (i.e., they would select those items that would increase their own motivation and competences), and know that the other students act rationally as well. Moreover, in games, it is assumed that the students are aware of their own predicted motivation from following each available strategy, as well as of the predicted motivation of the other group members from their choices. In order to suggest items to the group members, this information should become available to the game-theoretic (GT) group recommender, to guide decision support.

As stated in Section III-A, the predicted motivation $\hat{m}_{ij}$ for student *i* from item *j* is approximated with the Matrix Factorization technique [38]. The basic idea is to view the student-item motivation as a sparse matrix, for which we wish to predict the values of its empty cells, such that they would be consistent with the existing motivation values in the matrix. This is achieved by computing a low-rank approximation of the motivation matrix. As notational convention, bold small letters denote vectors, and bold capital letters denote matrices.

Let $\mathbf{M}$ be the matrix of size $|\mathrm{L}|\,\mathrm{x}\,|\mathrm{R}|$ that contains the motivation that the students get from the items. Each student l_i is associated with an *f*-dimensional factor vector $\boldsymbol{l}_i$, and similarly each item r_j with an *f*-dimensional factor vector $\boldsymbol{r}_j$. To get the predicted (approximated) motivation from an item r_j for student l_i, the inner product of the corresponding factor vectors is computed: $\hat{m}_{ij} = \mathbf{l}_i^T\mathbf{r}_j$. The resulting dot product captures the student's l_i overall motivation from the item r_j, and models this interaction. The major challenge is, then, to compute the mapping of each item and each student to the factor vectors, $\boldsymbol{r}_j$, $\boldsymbol{l}_i$, so that they accurately estimate the known motivation values without overfitting. The simplest approach to learn the factor vectors is to minimize the regularized squared error on the set of known motivation values:

$$\min \sum_{(l,r)\in K} (m_{ij} - \hat{m}_{ij})^2 + \lambda(\|l_i\|^2 + \|r_j\|^2)$$

where K is the set of $(l_i,\ r_j)$ pairs for which m_{ij} is known. The constant λ controls the extent of regularization and is usually determined by cross validation. To minimize this function and determine the factor vectors, Stochastic Gradient Descent [51] can be applied.

E. Student Grouping: Fuzzy C-Means (FCM)

A central topic in group RSs is the partition of the users into a number of groups, i.e., the group formation problem. The existing methods in educational group RSs promote shaping homogeneous groups of students (e.g., [29], [30]). However, as already stated in the motivation of the research section, having heterogeneous groups of students is considered as more beneficial in collaborative learning contexts, with respect to students' overall learning gain [33], [34]. Thus, both homogeneous and heterogeneous groups should be considered. Since the groups of students are not already known, clustering techniques are appropriate for detecting similarities (and dissimilarities) within the data.

A simple idea to end-up with homogeneous (heterogeneous) groups is to group the students based on the available individual ratings (the matrix containing the students' motivation from the items they have already rated), in such a way so that students with similar (dissimilar) ratings for the same items are in the same group. The FCM algorithm (the fuzzy version of the k-means algorithm) [52] was employed in this process. Unlike k-means, FCM allows data points to obtain fuzzy memberships to all clusters; in FCM, a student may belong to more than one group with a different probability. FCM works efficiently even with with small groups (i.e., 2–3 members) (e.g., [64], [65]).

Let $L = \{l_i\}$ be the set of students (i.e., data points) and $C = \{c_j\}$ be the set of clusters centers. Each student l_i is associated with an *f*-dimensional factor vector $\boldsymbol{l}_i$, and similarly each centroid c_j with an *f*-dimensional factor vector $\boldsymbol{c}_j$. For every cluster, a membership matrix $\mathbf{U}$ is created to represent the membership probabilities for every student. If u_{ij} is the membership probability of l_i in the cluster *j*, and *k* is the fuzziness index ($k > 1,\ k \in R$), the goal is to minimize (maximize) of the objective function

$$J_k = \sum_{i=1}^{n}\sum_{j=1}^{m} (u_{ij})^k \parallel l_i - c_j\|^2$$

where $\parallel * \parallel$ is the Euclidean distance between the $i\mathrm{th}$ student and the $j\mathrm{th}$ cluster center.

However, FCM takes as input the desired number of clusters and not the number of students per cluster, resulting in significant inequality in the sizes of the clusters (groups of students). During the clustering process, data points with extreme values tend to isolate, resulting in significantly less members in some of the created clusters than others. In order to automatically reform the clusters to become of equal size (preserving homogeneity), FCM's probability matrix U was utilized for exchanging members among the groups based on their highest probability of belonging to particular clusters, and according to a maximum number of members allowed in each cluster.

For the formation of mildly heterogeneous groups, after classifying students in homogeneous and heterogeneous groups, students can be randomly selected from different clusters, and regrouped in dyads and (or) triads.

IV. Experimental Evaluation

A. Participants and Experimental Setup

The proposed GT group-recommendation method was evaluated on a realistic setting with data from an empirical study with 102 students (55 girls [53.9%] and 47 boys [46.1%], aged 16 years old) from a European High School, in November 2017. The activity was about collaboratively writing simple functions in the Python programming language, using the recommended resources, as well as the lectures in the classroom. The collaboration between the students for writing the code for the requested functions, took place in face-to-face mode (the access to the resources, their evaluation, and the recommendation of the resources was in online mode). The students followed a semistructured design-thinking method, in terms that they were encouraged to write down their ideas on the solutions of the problem, to use pen and paper for preparing their code,

TABLE IV
DESCRIPTION OF GROUPS IN THE SECOND PHASE

Group ID	No. of students	Hom.	m. Het.	Het.	Rec. Method
T	26	3(x3)	3(x3)	4(x2)	GT
C1	26	3(x3)	3(x3)	4(x2)	AVG
C2	24	3(x3)	3(x3)	3(x2)	LM
C3	26	3(x3)	3(x3)	4(x2)	MP

and to test their code on the programming environment. All participants had no previous experience with programming, i.e., all of them were beginners (the course is introductory).

The experiment was conducted in three phases. During the first phase, 168 educational resources (i.e., worked examples, solved exercises, etc.), designed to motivate students and increase their interest in Python, were randomly assigned to the individuals. Each student had to study at least four, but not more than six items, within two days. After studing each item, the students had to rate it, assessing the overall motivation the item excited to the student, as explained in Section III-C.

The resulting dataset consisted of $|\mathrm{L}| = 102$ students, $|\mathrm{R}| = 148$ items, and $|\mathrm{M}^{\mathrm{L}}| = 687$ student-item ratings.

For the needs of the second phase, the students were arranged into four general, equivalent groups: one treatment ($T - 26$ students) and three control groups (*C1, C2, C3*-26, 24, 26 students, respectively). Each of these general groups was further partitioned in three types of subgroups with the FCM clustering method (described in Section III-E), and with respect to their members' previous ratings: 1) homogeneous (Hom), 2) heterogeneous (Het), and 3) mildly heterogeneous (m. Het). For each general group, $|\mathrm{G}| = 9$--10 subgroups were formed including 3 Hom., 3 m.Het. and 3 or 4 Het. (i.e., 39 subgroups in total), with $|\mathrm{g}|$ varying from two to three students per subgroup.

Next, one (or more) item(s) were delivered to each subgroup regularly (every two days) for two weeks, according to a recommendation strategy: the suggested GT method was applied on the subgroups of *T*, the Average method (*AVG*) was used on the subgroups of *C1*, the recommendations to the subgroups of *C2* were generated with *LM*, and the MP method provided the recommendations to subgroups of *C3*. *AVG, LM,* and *MP* are briefly demonstrated in [2], and a short description is in the following subsection (i.e., Section IV-B.2). Table IV summarizes the distribution of the students in groups and subgroups during the second phase of the activity, and maps the recommendation method to the corresponding group.

After studying the recommended items for two days, all group members had to rate them both individually, and as a team, using the same questionnaires as in the previous phase. Every second day, the $\mathbf{M}$ matrix, containing the real individual ratings, was updated. Another matrix, $\mathbf{V}$, containing the actual group ratings on the items was also constructed and updated. At the end of the second week, the student-item ratings were $|\mathrm{M}^{\mathrm{L}}| = 1464$, and the group-item ratings were $|\mathrm{M}^{\mathrm{G}}| = 232$. It should be noted that all subgroups of the control groups received one item per recommendation cycle, whereas the subgroups of the treatment group received up-to-three items per cycle. Throughout the experimental process, the items recommended to each subgroup should not have been previously rated by any of the subgroup members.

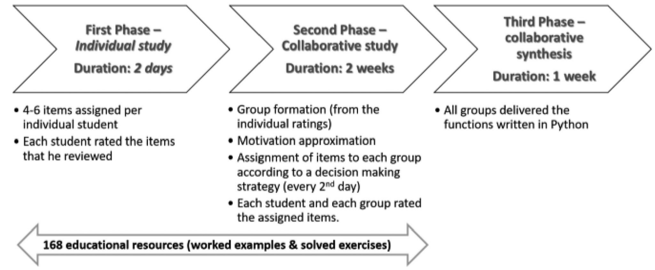

Fig. 2. Experimental activity process.

The third phase of the activity was about collaboratively writing simple functions in Python, using the resources recommended during the previous two phases, as well as their knowledge gained during the course lectures, within one week time. The activity cycle is illustrated and synopsized in Fig. 2.

Throughout the activity, the recommended items were delivered to the participants via a configured version of the Learning Analytics and Educational environment [53]. During all three phases, the items were available to the students to open and review, but not to download (in order to force the students' persistence); the students' time spent on viewing the recommended items, as well as the respective frequencies of reviewing them were being tracked (to measure groups' persistence). The students could assess the motivation that each item excited to them (either individually or as a team) only once, at the end of the respective recommendation cycle, when they were prompted to do so, but could not reconsider or modify their evaluation from that point on.

Finally, the collaborative problem-based learning activity was assessed by the course instructor according to the completeness and correctness of the final deliverable (the source code was produced with the Python 3.4.4 IDLE and submitted as a .py file), and was graded in a scale of [1, 20] as well.

B. Methods and Evaluation Metrics

1) Group Decision Strategies: As stated in the previous section, for each one of the control groups (i.e., *C1, C2,and C3*), the expected group motivation from an item was provided by a different group decision (aggregation) method, formulating how the corresponding subgroups of students reach to a consensus and come up with a decision about that particular item. Let *k* be the number of students in a group, $\hat{m}_{ij}$ the predicted motivation of student *i* from item *j*, the group motivation ratings were assigned according to the following strategies.

C1— Average (AVG): A consensus-based approach, where all group members jointly and equally make a decision. The group motivation equals the average motivation ratings across the group members: $M(k,j) = \frac{\sum_{i \in g} \hat{m}_{ij}}{k}$. In simple terms, *AVG* sets the average rating given by the group members to each item as the predicted rating of target group, and selects as recommendations those items that achieved the highest predicted ratings.

C2— LM: A borderline approach that targets to please the least happy member of the group, resulting the group to behave under a LM principle. In this case, the group motivation equals the minimum motivation among all group members: $M(k,j) = \min_{i \in g} \ \hat{m}_{ij}$. In other words, *LM* considers the rating of each item, assumes that the group's predicted rating on each item is the lowest value from the ratings given by all group members, and recommends these items. Thus, a group is as motivated as its least motivated member.

C3— MP: Another borderline group decision strategy satisfying the highest rating within the group. The motivation a group of students gets from an item equals to the maximum motivation within the group: $M(k,j) = \max_{i \in g} \ \hat{m}_{ij}$. Similarly to *LM, MP* takes under consideration the ratings of each item and next recommends the item with the maximum motivation among all group members. Thus, a group is as motivated as its most motivated member.

We decided to use these strategies as the control group methods because they have been previously identified in the Technology Enhanced Learning Group RSs literature that are the most commonly adopted strategies, which usually perform better. For example, in [27], Dwivedi and Bharadwaj compared the method they developed with the AVG and the LM methods. In another example [31], Zapata *et al.* evaluated and ranked 11 different aggregation strategies (identified in [26], p. 750), and it was found that the AVG method performs better and LM was the best performing boarderline approach. In addition, these methods have also been adopted by well-known commercial group RSs, such as PolyLens [23], which is using the LM method, and HappyMovie [24], which applies AVG. We decided to explore the MP method, as well, because we thought of maximizing motivation and hypothesized that this method could yield satisfactory results.

All solutions were implemented in MATLAB. Furthermore, the Gambit tool [54] was used to verify the correct identification of Nash Equilibria.

2) Evaluation Measures: Our proposed method targets at solving conflicts of interest by minimizing the prediction error of group motivation from the recommended educational resources (items). In the context of prediction accuracy estimation, the *Root Mean Square Error* (*RMSE*) is generally accepted as a good measure of precision, commonly used as an evaluation metric to compare prediction errors of different models for the same data. It measures the sample standard deviation of the difference between values approximated by an estimator and the values actually observed [55]. In our study, we explore the precision of our prediction with respect to motivation from the recommended items, as it is actually rated by *each* student, and by a given *group* of students. *RMSE* is computed as

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{n} \left(m_{kj} - \hat{m}_{kj}\right)^2}{n}}$$

where *n* is the number of items rated. Lower values indicate better predictions, and consequently, better decision strategy.

We also used the maximum *RMSE* for capturing the robustness of the RS, as it corresponds to the worst-case accuracy across *any* group. Lower *mRMSE* values indicate that all groups will receive good recommendations. This measure is computed as

$$m\mathrm{RMSE} = \max \sqrt{\frac{\sum_{j=1}^{n} \left(m_{kj} - \hat{m}_{kj}\right)^2}{n}}.$$

Furthermore, to measure the quality of the ranked list of recommended items delivered to groups of students, i.e., to evaluate its goodness, we used a measure from Information Retrieval, specifically crafted for ranking: the *Normalized Discounted Cumulative Gain* (*nDCG*), which assumes multiple levels of relevance [56]. In simple terms, *DCG* measures the gain of an item (i.e., the relevance score—if rating is missing, zero value is set) based on its position in the resulting list. The gain from the list is accumulated from top to bottom, and more relevant items are preferable to be on the top of the list (i.e., in our case, the most motivating). Thus, prior to accumulation, the scores are divided by the logarithm of the item's position, leading to a discount. *DCG* for a group of *k* students at position *N* (length of recommendation list), is computed as

$$\mathrm{DCG}_k@N = m_{kj_1} + \sum_{i=2}^{N} \frac{m_{kj_i}}{\log(i+1)}.$$

However, comparing *DCG*s between groups of students is not valid. As such, *nDCG* values are computed by arranging all items in an ideal order, and next dividing *DCG* by the ideal one (*IDCG*). Accordingly, *nDCG* is defined as

$$n\mathrm{DCG}_k@N = \frac{\mathrm{DCG}_k@N}{\mathrm{IDCG}_k@N}$$

where *IDCG* is the maximum possible *DCG*, and *nDCGk@N* getting values between 0 and 1, with 0 indicating the worst ranking and 1 representing the ideal ranking of items. In our study, due to limitations in available educational resources to be used as the recommendation items set, we only used short lists of up-to five items per group. Thus, we calculated *nDCG* with $\mathrm{N} = 3$ and $\mathrm{N} = 5$. We compared the effectiveness of both the *group* and *individual* recommendations when varying the aggregation method.

Finally, we measured the diversity of recommendation lists between different groups, by employing the *Hamming Distance (HD)* [57] metric. *HD* estimates if the recommendations to all groups make full use of all items, leaving only a few items without being recommended. If $Q_{g,g*}$ is the "overlapped" number of items recommended to both groups *g* and *g*∗, respectively, then the *HD* between group *g* and group *g*∗, is defined as

$$\mathrm{HD}(g, g*) = 1 - \tfrac{Q_{g,g*}}{|z|}$$

where z is the length of the recommendation list. High *HD* means high diversity, making full use of all items and leaving out of recommendation only a few items; a highly personalized recommendation list should have higher *HD* to other lists.

TABLE V
(A) METRICS FOR HOMOGENEOUS GROUPS, (B) METRICS FOR MILDLY HETEROGENEOUS GROUPS, (C) METRICS FOR HETEROGENEOUS GROUPS

(A)

	RMSE	*mRMSE*	*nDCG@3*	*nDCG@5*	*HD*
GT	**0.342**	0.475	0.954	**0.961**	**0.871**
AVG	0.351	**0.448**	**0.959**	0.958	0.834
LM	0.386	0.647	0.883	0.884	0.715
MP	0.392	0.724	0.877	0.875	0.686

(B)

	RMSE	*mRMSE*	*nDCG@3*	*nDCG@5*	*HD*
GT	**0.366**	**0.515**	**0.949**	**0.950**	**0.843**
AVG	0.564	0.738	0.912	0.913	0.754
LM	0.785	1.233	0.806	0.805	0.622
MP	0.800	1.839	0.774	0.772	0.604

(C)

	RMSE	*mRMSE*	*nDCG@3*	*nDCG@5*	*HD*
GT	**0.416**	**0.524**	**0.942**	**0.940**	**0.832**
AVG	0.773	0.958	0.852	0.851	0.674
LM	1.092	1.645	0.716	0.716	0.504
MP	1.521	2.132	0.667	0.668	0.498

V. RESULTS

A. *Prediction Accuracy, Effectiveness, and Diversity of Recommendations and Students' Persistence for Groups*

Tables V(a), (b), and (c) demonstrate the results for the evaluation metrics (average values) for all decision support strategies compared in this study, i.e., the currently proposed GT method applied on the treatment group, and the *AVG*, *LM*, and *MP* methods applied on each one of the control groups, for homogeneous (high inner member similarity), mildly heterogeneous (medium inner member similarity), as well as heterogeneous (low inner member similarity) synthesis of the subgroups, respectively. The subgroups sizes was firm, varying from two to three students, as explained in Section IV-A.

According to these results, all decision support methods achieve low approximation error in prediction of motivation ratings for the homogeneous students' groups, as expected. On the contrary, for mildly heterogeneous groups, accuracy is high for the GT and satisfactory for the AVG method, but the prediction error significantly increases when the aggregation strategy is LM or MP. For heterogeneous groups, the approximation error in prediction of motivation is low only when the recommendation strategy is the GT, whereas it is high for all the other cases.

Furthermore, the group recommendation effectiveness tends to decrease only for the heterogeneous subgroups. Fig. 3 Illustrates the average goodness of the ranked list of recommended items delivered to the subgroups of students when the top ranked items are 5 (nDCG@5) and when the top ranked items are 3 (nDCG@3), according to the inner similarity of the subgroups, and by considering the decision support strategy.

Diversity in recommendations, reflected in the HD values, indicates that the recommendation to all groups—the

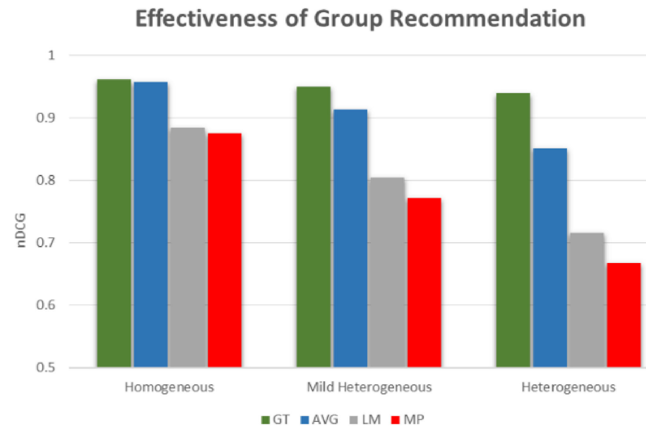

Fig. 3. (Average) Effectiveness of group recommendations with respect to the aggregation strategy and the group inner similarity.

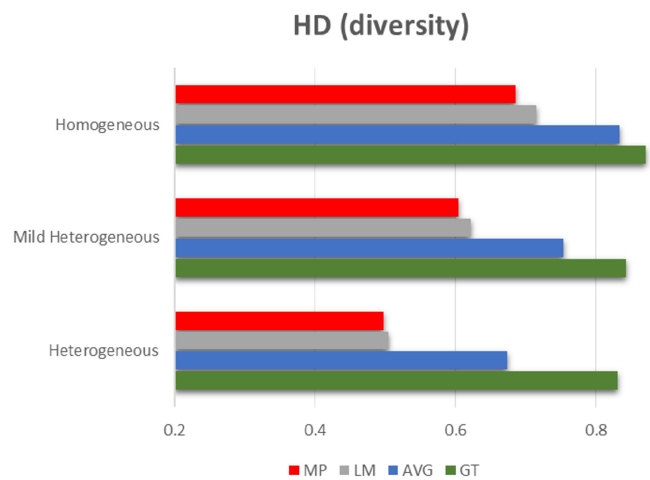

Fig. 4. Diversity of group recommendations with respect to the aggregation strategy and the group inner similarity.

homogeneous and the (mildly) heterogeneous—make sufficient use of all items and few items will be left without being recommended, when the recommendation method is the proposed GT method. Fig. 4 illustrates the diversity in recommendation according to the different decision support strategy for either the homogeneous or the (mildly) heterogeneous groups.

Finally, Fig. 5 illustrates the groups' average time spent on reviewing the items (based on the total time spent to review the items and the respective frequencies of reviewing) throughout the second and third phases of the activity, for all decision support strategies compared in this study. As stated in Section IV-A, this measure is indicative of students' actual engagement with the items, and codes their persistence.

B. *Effectiveness of Recommendations and Students' Conformity for Individual Students Within Groups*

Furthermore, in order to understand *when* the group recommendations are better or worse (ranked) for each individual within the subgroups, we measured the difference between the effectiveness of the individual and the group recommendations'

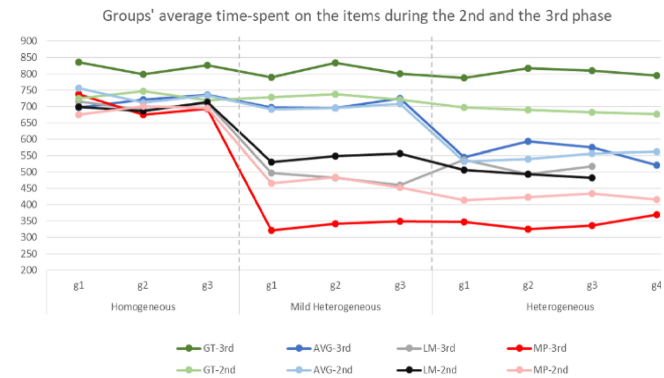

Fig. 5. Groups' actual engagement with the items during the collaborative phases of the activity, with respect to the aggregation strategy and the group similarity.

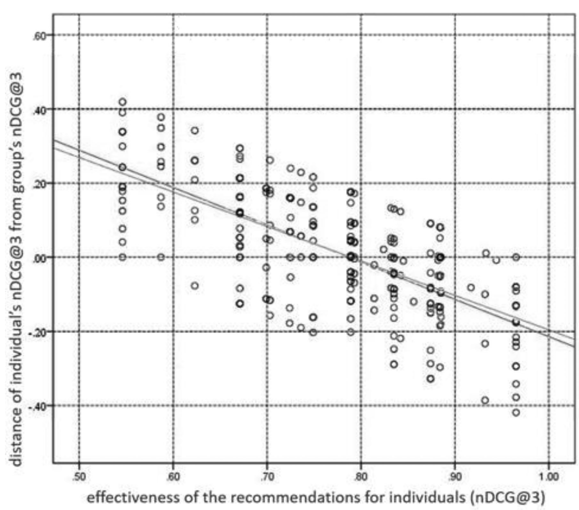

Fig. 6. Distance of group motivation with respect to individual motivation.

lists. This difference is indicative of the individuals' degree of conformity, regarding the adjustment of their motivation from the recommendation with respect to the motivation of the group they are members of. A positive difference means that the group recommendations are better ranked than the individual recommendations. Fig. 6 shows a scatter plot where each student, in a group, is represented by a point, for the two better performing methods, i.e., the GT and the AVG method.

Here, the x-axis measures nDCG@3 for the individual recommendation list, while the y-axis shows the distance of the individual's from the respective group's nDCG of this group recommendation list for the same student. In this figure, the green trendline corresponds to the GT method, whereas the red trendline corresponds to the AVG method, respectively.

Please note that during the two weeks of experimentation, each student and each group received recommendations every second day, resulting to a total of more than one recommendations, and, hence, a student may be represented by several points.

C. *Effect of Recommendations on Group Learning Performance*

Regarding the effect of the recommendation strategy on the groups' performance in the collaborative learning activity, as it

TABLE VI
EFFECT OF RECOMMENDATION STRATEGY ON PERFORMANCE

	Mean (SD)	Mean (SD)	Mean Diff.	t-value	p-value	Cohen's d
GT – AVG	17.00 (1.054)	15.20 (2.150)	1.80	2.377*	0.029	1.063
GT – LM		14.78 (1.922)	2.22	3.171*	0.006	1.432
GT – MP		13.90 (1.524)	3.10	5.291*	0.000	2.365
AVG – LM	15.20 (2.150)	14.78 (1.922)	0.42	0.449	0.659	0.205
AVG – MP		13.90 (1.524)	1.30	0.208	0.136	0.697
LM – MP	14.78 (1.922)	13.90 (1.524)	0.88	1.109	0.283	0.507

$p<0.05$ (2-tailed).

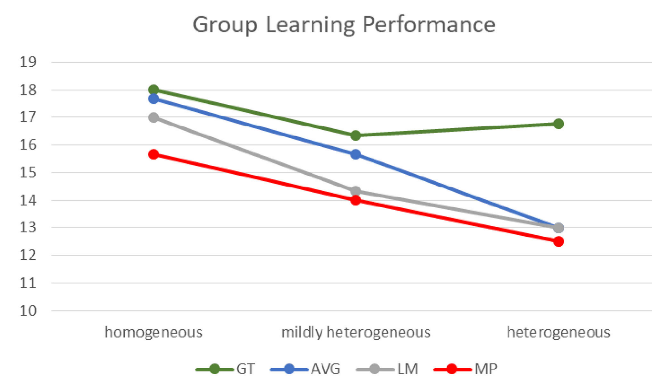

Fig. 7. Means of learning performance per recommendation strategy per group homogeneity type.

is reflected on their grades, *independent sample t-tests* were conducted and the *Cohen's d effect size* was calculated. Results indicate that taking the recommended resources has a differentiated impact on performance depending on the groups' achievement level in the collaborative activity. Table VI compares the average learning performance for all four recommendation strategies for all types of group inner homogeneity level.

Finally, Fig. 7 plots the mean scores per group inner homogeneity type per recommendation strategy.

VI. DISCUSSION

Recommending educational resources to groups of students, targeting at optimizing all students' motivation, both individually and as a group, is a complicated task. The core issue is to determine how a group of students reaches to a consensus about their degree of appreciation for each item, in such a way that reflects the self-determined interests and motivation *of each and all* group members. This study focuses on addressing diversity in group members' self-motivated considerations regarding the usefulness of the resources, as well as in their behavioral intention to finally use the resources (conflict of interest).

The review of related research identified four types of drawbacks related to: 1) the promotion of high group inner similarity, 2) the effectiveness of aggregation strategies, 3) the number of recommended resources, and 4) the omission of the individual members' conformity degree [27]–[31]. Moreover, none of these approaches—to the best of our knowledge—explored the actual engagement of the group members with the recommended items, in a learning analytics fashion, explaining the students' persistence, and reasoning the final learning performance in the collaborative activity. As such, the emerging research questions were as follows.

RQ1: Can we accurately and efficiently recommend sequences of educational resources to homogeneous and heterogeneous groups of students, with respect to both the individuals' and the group's motivation, and intention to use the resources?

RQ2: What is the impact of a recommendation on individual students' persistence as well as on the groups' learning performance in the collaborative problem-solving activity?

In order to address the abovementioned issues and answer the research questions, this study suggested and evaluated a noncooperative GT perspective for solving conflict of interest between the group members and guiding the recommendation process.

For the empirical evaluation of the approach in a realistic setting, data were collected during a collaborative problem-oriented learning activity with 102 students from a European High School. The objectives were twofold: 1) to compare the accuracy and the effectiveness of ranked lists of recommended items delivered to groups of students by the suggested method to other state-of-the-art decision support methods, with respect to the individual's motivation from the recommended items, and 2) to explore the impact of recommendation on individual members' level of conformity, as well as on the overall groups' persistence and learning performance. The following novel facts and important observations have risen.

A. Accuracy and Effectiveness of the Recommendation for Different Levels of Group Inner Similarity

First, from Table V it becomes apparent that all decision support methods achieve low approximation error in prediction of motivation ratings for the homogeneous students' subgroups. However, the proposed GT strategy minimizes the prediction error of the subgroup motivation ratings, as, by far, it scores the lowest RMSE values for all categories of inner subgroup similarity. Especially for the highly heterogeneous subgroups, the other aggregation methods combine potentially conflicting rankings that could create a group recommendation, which might not be motivating for the group members; accuracy is high for the GT method ($RMSE = 0.416$), but the prediction error significantly increases when the aggregation strategy is AVG ($RMSE = 0.773$), LM ($RMSE = 1.092$), or MP ($RMSE = 1.521$). As such, in this case, the GT decision strategy resolves sufficiently the conflict of interest and delivers the most appropriate items to the students.

Still, mRMSE demonstrates some variance in the prediction error across subgroups. In particular, for the homogeneous subgroups, the GT method did not have the lowest prediction error ($mRMSE = 0.475$); in this case, it turns out that the AVG strategy was a better approach, although only slightly ($mRMSE = 0.448$). Yet, this was an expected finding, since the average method works well in most cases of homogeneous groups [28]–[30].

Second, from the same table of results, one can observe that the suggested GT method has a good overall performance (i.e., the nDCG values reflecting the effectiveness of ranked list or recommendations), although not always the best; in one case of homogeneous groups, the AVG method provided slightly more effective recommendations ($nDCG@3 = 0.959$) compared to the list of GT ($nDCG@3 = 0.954$). However, it is important to notice that, compared to the other methods, the performance of the proposed GT seems to be stable and robust, regardless of the inner subgroup similarity, targeting ranking quality and demonstrating only small variations. From the evaluation results, it was found that nDCG for the GT method is close to 1.0 (higher than 0.9) in all cases of subgroup homogeneity, whereas the respective values for the other methods decrease as the inner group similarity decreases.

Third, another finding concerns the plurality of the recommended lists of items (diversity of items). It is very possible that the deficiency in capturing the whole range of students' needs and preferences could lead to poor motivation. Improving the diversity of recommendation results is expected to increase within group motivation from the recommended resources. As seen in Table V, the values of the HD metric reflect that the personalization of recommendation is better for homogeneous groups, regardless of the method employed, whereas the GT method provides satisfactory personalization even for heterogeneous groups ($HD = 0.832$) compared to the other methods (*HD varying from 0.498 to 0. 674*). In addition, this finding implies that the GT method makes sufficient use of all items and only few items will be left without being recommended.

B. Impact of Recommendation on Conformity, Persistence, and Learning Performance

The second research question concerned the impact of the recommendation on the individuals' gain as well as on the overall groups' performance. For this purpose, three indices were employed: 1) the individuals' conformity degree, expressed as the distance between the individual nDCG and the respective group's nDCG, 2) the actual aggregated time spent on the recommended items and the frequencies of reviewing them, as measures of students' persistence, in a learning analytics fashion, and 3) the groups' grades (scores) in the collaborative activity.

First, in order to understand how much the individuals adjusted their personal evaluation of motivation from the recommended items compared to the group's they belong to (conformity degree), we plotted the best fit line through the points in the scatter plot (in Fig. 6). One can observe that when the employed method is the GT, the gradient of the curve is smaller compared to the next best performing method, i.e., the AVG. This means that the recommendations are highly ranked for the individuals as well as for the group. In other words, the individuals within the groups do not have to highly adjust their personal consideration about their motivation gained from the recommended items, to be approved by the other group members. The measure employed, i.e., the distance between the individual nDCG and the respective group's nDCG, is a topic that deserves further analysis and could be explored as a measure of the individual's degree of conformity.

Second, as seen from the curves in Fig. 5, the students' engagement with the resources during both collaborative phases of the activity, and regardless of the inner group similarity, was stable and high when the recommendations were generated with the GT method. The same fact is also true for the mildly heterogeneous groups when the recommendation method is AVG, as well. When the inner group member similarity is high (homogeneous groups), all methods deliver items that motivate students to actually participate in the problem-oriented collaborative activity. Yet, for heterogeneous groups, the students' disengagement from the resources successively increases for the AVG, LM, and MP methods. The disengagement is even higher for these methods during the third phase, when the groups had to complete the collaborative assignment by using the recommended resources. This fact is illustrated on the average time spent on the resources and on the frequency of reviewing them to complete the task.

Finally, the independent samples t-tests for comparing the differences in groups' final grades with respect to the recommendation method verified that the resources delivered to the groups affected their learning performance, as expected.

VII. Conclusion

In this study, we aimed at recommending educational resources to groups of students with diverse levels of inner member similarity. The goal was to decide upon those resources that would best support *each* and *all* group members to efficiently complete the assigned collaborative activity.

Inspired from [3], we argue that *Game Theory* could efficiently solve "conflicts of interest" between the group members [4] and guide the recommendation of educational resources. Game theory is about social situations, providing solid recommendations to the players regarding their own optimal strategy, as well as administering an external observer that predicts the outcome of interactions (i.e., in our approach, the decision support system). The proposed solution models the recommendation strategy as a problem of finding the NE, i.e., a state in which no student can be benefited more in terms of further improving their own motivation by unilaterally deviating from the NE. However, the best collective result does not always come from each individuals following their own interest, but rather from reaching the group's consensus; whereas an NE does not correspond to a socially optimal outcome, a Pareto optimal equilibrium describes a social optimum in the sense that no individual player can improve their payoff without making at least one other player worse off. Pareto optimality is not a solution concept, but is used to evaluate the overall gain.

To this end, we developed a mathematical formulation (i.e., Algorithm 1) for group-recommendation of educational resources as a noncooperative game, as well as an architecture for building such group RSs (i.e., Fig. 1).

From the evaluation of the approach with a realistic dataset, results revealed a tendency that the accuracy of the predicted group motivation, the goodness of the ranked list of recommendations, and the problem-solving performance for the treatment group were significantly higher compared to other state-of-the-art methods. The diversity of the items in the recommendation, indicating a satisfactory personalization even for heterogeneous groups, was higher with the GT method, as well.

We also explicitly compared the differences in individual evaluation of motivation from the recommendation, to the group's perception; we introduced the distance between the individual nDCG and the respective group's nDCG, as a measure of students' degree of conformity. Additional research is required.

Moreover, we adopted a learning analytics view point to explore the groups' actual engagement with the recommended resources and evaluated the effect of their persistence on the final learning performance. It was confirmed that the more motivating the recommended items, and the higher the students' persistence, the better the learning outcome.

However, there are some limitations. First, the samples of the 168 educational resources and 102 students considered in the evaluation process are small and potentially biased; bigger datasets should be analyzed. Second, we investigated only groups of two to three students; the behavior of GT with larger groups of students (e.g., 4 to 5 members) should be explored as well. Finally, we assumed that the group formation method used in this study would not raise issues of uncertainty; other methods for group formation should be explored as well.

Another limitation of the present analysis—yet, not directly related to the suggested GT group recommendation approach—is that during the experimental evaluation of the method, we did not provide a mechanism to prevent students from downloading the material (educational resources) and studying it in offline mode. This behavior could potentially impact (either negatively or positively—we cannot say without exploring it) the learners' persistence/efforts/motivation. However, the option of downloading the material is common for all groups (both the treatment and the control groups), and thus, it cannot be considered as biased only for the treatment group.

It also should be noted that all subgroups of the control groups received one item per recommendation cycle, whereas the subgroups of the experimental group received up to three items per cycle. This is a significant difference of the method, compared to the other (control) methods, which deliver one recommendation item per cycle. In a sense, delivering a sequence of items could be perceived as a different "treatment" of the experimental group, which is expected to affect the recommendation performance.

Elaborating more on that aspect, and generally speaking, the issue of recommending a sequence of items that is pleasing *as a whole*, instead of focusing on the generation of a single recommendation, has been a major topic in the RS literature [26], [58]–[60]. The combination of recommending to a group *and* recommending a sequence has been acknowledged as very interesting [26], as it may allow keeping all individuals in the group satisfied, by compensating for items a particular user dislikes with other items in the sequence which they do like [26], [58]. Sequences of items are in a sense more "flexible" recommendations, because sequences not only afford the RS more chances to make accurate recommendations, but also to mix familiar and unfamiliar items [59].

There are cases in which recommending sequences of items per cycle is important. For example, in the scenarios of MusicFX [21] and PolyLens [23], it is sufficient to recommend single items: people normally only see one movie per evening, while radio stations can play forever [2]. However, in learning processes, it is very likely that the individual students might want to access multiple resources to meet their learning goals/interest and intrinsic motivation. Similarly, in group learning contexts, it is very possible that, in (mildly) heterogeneous groups, most of the learners prefer to access different learning items. In TEL RSs, it is a quite common task to recommend sequences of items to individuals [59]. However, in TEL group RSs, the topic has not been yet sufficiently explored—to the best of our knowledge. Furthermore, the other aggregation methods fail to deliver such sequences of items per recommendation cycle. A lot more research is needed on algorithms and user interfaces for producing coherent sequences of recommendations [26].

Moreover, although it might seems that the sequence of items better fits the group's motivational excitement, resulting in improved performance for the (mildly) heterogeneous groups, however, recommending more items also implies more effort on viewing/reviewing and evaluating them. We cannot say at the moment, which exactly is the effect of delivering sequences of items instead of single items, and further analyses are required toward addressing this issue. In this paper, we did not analyze the difference between the GT groups who received single items and the GT groups that received sequences of items per recommendation cycle. This is within our future work plans.

In addition, a number of other challenges for future work have emerged. For example, more sophisticated measures of motivation and persistence could be applied (e.g., incorporating the students' affective states, perceived enjoyment, challenge). The learning analytics research could contribute toward this direction. Yet, another challenging issue is focusing on the transparency of the group recommendation: showing each individual's payoff and, eventually, how motivated the other group members are, could improve the particular student's understanding of the recommendation process, and perhaps make it easier to accept the educational resources that initially he/she did not like. Finally, one more open topic is the use of *cooperative games* (instead of noncooperative, presented in this paper), to model the problem of group recommendation. This alternative approach needs to be further explored.

To conclude, the core contribution of this study is that the proposed GT solution demonstrates a socially and individually optimum group recommendation method, based on the motivational effect of the resources and the students' intention to use them, and yields statistically significant results even for highly heterogeneous groups of students.

REFERENCES

[1] S. Berkovsky and J. Freyne, "Group-based recipe recommendations: Analysis of data aggregation strategies," in *Proc. 4th ACM Conf. Recommender Syst.*, 2010, pp. 111–118.

[2] J. Masthoff, "Group recommender systems: Combining individual models," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Boston, MA, USA: Springer, 2011, pp. 677–702.

[3] L. A. M. C. Carvalho and H. T. Macedo, "Users' satisfaction in recommendation systems for groups: An approach based on noncooperative games," in *Proc. 22nd Int. Conf. World Wide Web*, 2013, pp. 951–958.

[4] J. Nash, "Non-cooperative games," *Ann. Math.*, vol. 54, no. 2, pp. 286–295, 1951.

[5] R. B. Myerson, "Negotiation in games: A theoretical," Essays in Honor of Kenneth J. Arrow: Uncertainty, Information, and Communication, vol. 3, Sep. 1986.

[6] P. Resnick and H. R. Varian, "Recommender systems," *Commun. ACM*, vol. 40, no. 3, pp. 56–58, Mar. 1997.

[7] J. Bobadilla, F. Ortega, A. Hernando, and A. GutiéRrez, "Recommender systems survey," *Know.-Based Syst.*, vol. 46, pp. 109–132, Jul. 2013.

[8] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, "Recommender system application developments," *Decis. Support Syst.*, vol. 74, no. C, pp. 12–32, Jun. 2015.

[9] D. H. Park, H. K. Kim, I. Y. Choi, and J. K. Kim, "A literature review and classification of recommender systems research," *Expert Syst. Appl.*, vol. 39, no. 11, pp. 10059–10072, Sep. 2012.

[10] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proc. 10th Int. Conf. World Wide Web*, 2001, pp. 285–295.

[11] M. J. Pazzani and D. Billsus, *The Adaptive Web*. P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds., Berlin, Germany: Springer-Verlag, 2007, pp. 325–341.

[12] R. Burke, "Hybrid recommender systems: Survey and experiments," *User Model. User-adapt. Interact.*, vol. 12, no. 4, pp. 331–370, Nov. 2002.

[13] R. R. Yager, "Fuzzy logic methods in recommender systems," *Fuzzy Sets Syst.*, vol. 136, no. 2, pp. 133–149, Jun. 2003.

[14] J. He and W. W. Chu, "A social network-based recommender system (SNRS)," in *Data Mining for Social Network Data*, N. Memon, J. J. Xu, D. L. Hicks, and H. Chen, Eds. Boston, MA, USA: Springer, 2010, pp. 47–74.

[15] G. Adomavicius and A. Tuzhilin, "Context-aware recommender systems," in *Proc. ACM Conf. Recommender Syst.*, 2008, pp. 335–336.

[16] K. McCarthy, M. Salamó, L. Coyle, L. McGinty, B. Smyth, and P. Nixon, "Group recommender systems: A critiquing based approach," in *Proc. 11th Int. Conf. Intell. User Interfaces*, 2006, pp. 267–269.

[17] S. Amer-Yahia, S. B. Roy, A. Chawlat, G. Das, and C. Yu, "Group recommendation: Semantics and efficiency," *Proc. VLDB Endow.*, vol. 2, no. 1, pp. 754–765, Aug. 2009.

[18] D. M. Pennock, E. Horvitz, and C. L. Giles, "Social choice theory and recommender systems: Analysis of the axiomatic foundations of collaborative filtering," in *Proc. 17th Nat. Conf. Artif. Intell. 12th Conf. Innovative Appl. Artif. Intell.*, 2000, pp. 729–734.

[19] P. C. Fishburn, *The Theory of Social Choice*. Princeton, NJ, USA: Princeton Univ. Press, 1973.

[20] J. Masthoff and A. Gatt, "In pursuit of satisfaction and the prevention of embarrassment: affective state in group recommender systems," *User Model. User-Adapt. Interact.*, vol. 16, no. 3, pp. 281–319, Sep. 2006.

[21] J. F. McCarthy and T. D. Anagnost, "MusicFX: An arbiter of group preferences for computer supported collaborative workouts," in *Proc. ACM Conf. Comput. Supported Cooperative Work*, 1998, pp. 363–372.

[22] L. Ardissono, A. Goy, G. Petrone, M. Segnan, and P. Torasso, "INTRIGUE: Personalized recommendation of tourist attractions for desktop and hand held devices," *Appl. Artif. Intell.*, vol. 17, no. 8, pp. 687–714, 2003.

[23] M. O'Connor, D. Cosley, J. A. Konstan, and J. Riedl, "PolyLens: A recommender system for groups of users," in *Proc. 7th Eur. Conf. Comput. Supported Cooperative Work*, 2001, pp. 199–218.

[24] L. Quijano-Sanchez, J. A. Recio-Garcia, B. Diaz-Agudo, and G. Jimenez-Diaz, "Social factors in group recommender systems," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 1, pp. 8:1–8:30, Feb. 2013.

[25] M. Gartrell, X. Xing, Q. Lv, A. Beach, R. Han, S. Mishra, and K. Seada, "Enhancing group recommendation by incorporating social relationship interactions," in *Proc. 16th ACM Int. Conf. Supporting Group Work*, 2010, pp. 97–106.

[26] J. Masthoff, "Group recommender systems: Aggregation, satisfaction and group attributes," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, and B. Shapira, Eds. Boston, MA, USA: Springer, 2015, pp. 743–776.

[27] P. Dwivedi and K. K. Bharadwaj, "e-Learning recommender system for a group of learners based on the unified learner profile approach," *Expert Syst.*, vol. 32, no. 2, pp. 264–276, 2015.
[28] P. Rodríguez, M. Giraldo, V. Tabares, N. Duque, and D. Ovalle, "Recommendation system of educational resources for a student group," in *Proc. Highlights Practical Appl. Scalable Multi-Agent SystPAAMS Collection: Int. Workshops*, 2016, pp. 419–427.
[29] M. Kompan and M. Bielikova, "Enhancing existing e-learning systems by single and group recommendations," *Int. J. Contin. Eng. Educ. Life Long Learn.*, vol. 26, no. 4, pp. 386–404, 2016.
[30] D. Yanhui, W. Dequan, Z. Yongxin, and L. Lin, "A group recommender system for online course study," in *Proc. 7th Int. Conf. Inf. Technol. Med. Educ.*, 2015, pp. 318–320.
[31] A. Zapata, V. H. Menéndez, M. E. Prieto, and C. Romero, "Evaluation and selection of group recommendation strategies for collaborative searching of learning objects," *Int. J. Hum. Comput. Stud.*, vol. 76, no. Supplement C, pp. 22–39, 2015.
[32] S. E. Asch, "Effects of group pressure upon the modification and distortion of judgments," in *Groups, Leadership and Men; Research in Human Relations*. Oxford, U.K.: Carnegie Press, 1951, pp. 177–190.
[33] S. Graf and R. Bekele, "Forming heterogeneous groups for intelligent collaborative learning systems with ant colony optimization," in *Proc. 8th Int. Conf. Intell. Tutoring Syst.*, 2006, pp. 217–226.
[34] R. E. Slavin, "Developmental and motivational perspectives on cooperative learning: A reconciliation," *Child Dev.*, vol. 58, no. 5, pp. 1161–1167, 1987.
[35] D. L. Medin, R. L. Goldstone, and D. Gentner, "Similarity involving attributes and relations: Judgments of similarity and difference are not inverses," *Psychol. Sci.*, vol. 1, no. 1, pp. 64–69, 1990.
[36] T. Mussweiler, "Comparison processes in social judgment: mechanisms and consequences," *Psychol. Rev.*, vol. 110, no. 3, pp. 472–489, Jul. 2003.
[37] Z. Yu, X. Zhou, Y. Hao, and J. Gu, "TV program recommendation for multiple viewers based on user profile merging," *User Model. User-Adapt. Interact.*, vol. 16, no. 1, pp. 63–82, Mar. 2006.
[38] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Comput. (Long. Beach. Calif).*, vol. 42, no. 8, pp. 30–37, Aug. 2009.
[39] P. R. Pintrich and D. H. Schunk, *Motivation in Education : Theory, Research, and Applications*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.
[40] J. Brophy, *Motivating Students to Learn*, 2nd ed. Mahwah, NJ, USA: Lawrence Erlbaum Associates Publishers, 2004.
[41] A. Wigfield and J. S. Eccles, "Expectancy–value theory of achievement motivation," *Contemp. Educ. Psychol.*, vol. 25, no. 1, pp. 68–81, 2000.
[42] B. Weiner, "An attributional theory of achievement motivation and emotion," *Psychol. Rev.*, vol. 92, no. 4, pp. 548–573, 1985.
[43] C. S. Dweck, "Motivational processes affecting learning," *Am. Psychol.*, vol. 41, no. 10, pp. 1040–1048, Oct. 1986.
[44] E. L. Deci and R. M. Ryan, Eds., *Handbook of Self-Determination Research*. Rochester, NY, USA: University of Rochester Press, 2002.
[45] D. H. Lim, "Cross cultural differences in online learning motivation," *Educ. Media Int.*, vol. 41, no. 2, pp. 163–175, 2004.
[46] R. M. Ryan and E. L. Deci, "Intrinsic and extrinsic motivations: Classic definitions and new directions," *Contemp. Educ. Psychol.*, vol. 25, no. 1, pp. 54–67, Jan. 2000.
[47] Y.-M. Huang, "Exploring students' acceptance of team messaging services: The roles of social presence and motivation," *Brit. J. Educ. Technol.*, vol. 48, no. 4, pp. 1047–1061, 2017.
[48] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quart*., vol. 13, no. 3, pp. 319–340, Sep. 1989.
[49] V. Terzis and A. A. Economides, "The acceptance and use of computer based assessment," *Comput. Educ.*, vol. 56, no. 4, pp. 1032–1044, 2011.
[50] M. L. Maehr and H. A. Meyer, "Understanding motivation and schooling: Where we've been, where we are, and where we need to go," *Educ. Psychol. Rev.*, vol. 9, no. 4, pp. 371–409, Dec. 1997.
[51] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proc. 19th Int. Conf. Comput. Statist.*, 2010, pp. 177–186.
[52] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. Norwell, MA, USA: Kluwer, 1981.
[53] Z. K. Papamitsiou and A. A. Economides, "Towards the alignment of computer-based assessment outcome with learning goals: The LAERS architecture," in *Proc. IEEE Conf. e-Learn., e-Manage. e-Services*, 2013, pp. 13–17.
[54] R. D. Mckelvey, A. M. Mclennan, and T. L. Turocy, "Gambit: Software tools for game theory," Version 16.0.1, 2006. [Online]. Available: http://www.gambit-project.org
[55] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecasting*, vol. 22, no. 4, pp. 679–688, 2006.
[56] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques," *ACM Trans. Inf. Syst.*, vol. 20, no. 4, pp. 422–446, Oct. 2002.
[57] T. Zhou, L.-L. Jiang, R.-Q. Su, and Y.-C. Zhang, "Effect of initial configuration on network-based recommendation," *Europhysics Lett.*, vol. 81, Mar. 2008, Art. no. 58004.
[58] A. Jameson and B. Smyth, "Recommendation to groups," in *The Adaptive Web: Methods and Strategies of Web Personalization*, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Berlin, Germany: Springer, 2007, pp. 596–627.
[59] N. Tintarev, C., Lofi, and C. C. S. Liem, "Sequences of diverse song recommendations: An exploratory study in a commercial system," in *Proc. 25th Conf. User Model. Adaptation Personalization*, 2017, pp. 391–392.
[60] H. Drachsler, K. Verbert, O. C. Santos, and N. Manouselis, "Panorama of recommender systems to support learning," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, Eds. Boston, MA, USA: Springer, 2005, pp. 421–451.
[61] I. Ajzen, "From intentions to actions: A theory of planned behavior," in *Action Control: From Cognition to Behavior*, J. Kuhl and J. Beckmann, Eds. Berlin, Germany: Springer, 1985, pp. 11–39.
[62] A. Bandura, *Social Learning Theory*. Oxford, U.K.: Prentice-Hall, 1977.
[63] U. Hanani, B. Shapira, and P. Shoval, "Information filtering: Overview of issues, research and systems," *User Model. User-Adapt. Interact.*, vol. 11, no. 3, pp. 203–259, Aug. 2001.
[64] C. E. Christodoulopoulos and K. A. Papanikolaou, "A group formation tool in an e-learning context," in *Proc. 19th IEEE Int. Conf. Tools Artif. Intell.*, 2007, pp. 117–123.
[65] D. Nurjanah, K. Dewanto, and F. D. Sari, "Homogeneous group formation in collaborative learning using fuzzy C-means," in *Proc. IEEE 6th Int. Conf. Teaching Assessment Learn. Eng.*, 2017, pp. 74–79.

Zacharoula Papamitsiou (Member, IEEE) received the M.Sc. degree in educational technologies from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2011. She is currently working toward the Ph.D. degree in learning analytics and educational data mining for adaptive assessment at the interdepartmental program of postgraduate studies in information systems from the University of Macedonia, Thessaloniki, Greece.

She has published more than 15 peer-reviewed papers. Her current research interests include learning analytics, educational data mining, data-driven decision making, educational recommender systems, and adaptive feedback and assessment.

Dr. Papamitsiou has received the Martin Wolpers Award for the research project of the most promising young researcher (2017). She is a Reviewer in more than five high ranked journals. She is a Member of the IEEE and the IEEE Computer Society, Society for Learning Analytics Research, and the ACM.

Anastasios A. Economides received the M.Sc. and Ph.D. degrees in computer engineering, from the University of Southern California, Los Angeles, CA, USA, in 1987 and 1990, respectively.

He is with the University of Macedonia (UoM), Thessaloniki, Greece. He is a Member of the Quality Assurance Unit of the UoM and Deputy representative of UoM at EL.ID.E.K. (the official administration for the National/Greek Research Policy). He is the Director of of the Smart & Mobile Interactive Learning Environments lab. He has published 2 books, more than 100 peer-reviewed journal papers, more than 150 peer-reviewed conference papers and, more than 20 peer-reviewed chapter books. He has received around 3500 citations. His current research interests include e-learning, networking technologies and techno-economics, and competition in telecommunication networks. Also, he was the keynote speaker in international conferences, in the program committees of dozens of conferences and in the editorial committee boards of many journals.