# Recommendation of educational resources to groups: a game-theoretic approach

Zacharoula Papamitsiou, Anastasios A. Economides
Interdepartmental Program of Postgraduate Studies in Information Systems
University of Macedonia
Thessaloniki, Greece
papamits@uom.edu.gr, economid@uom.gr

***Abstract*— In collaborative learning contexts, it is necessary to recommend educational resources to groups of students instead of individuals. However, this task is not trivial, because students in a group may not be fulfilled by the same items, yet wish to meet their own expectations. Existing approaches either merge individual profiles and recommend items accordingly, or fuse the lists of individual recommendations. Both perspectives achieve low quality performance and goodness of recommendation for majority of students in heterogeneous groups. This paper follows a game-theoretic approach for solving conflict of interest among students and recommending resources to both homogeneous and heterogeneous groups in collaborative learning contexts. The group members are the players, the resources comprise the set of possible actions, and selecting those items that will maximize all students' satisfaction – both individually and as a whole – is a problem of finding the Nash Equilibrium. During the empirical evaluation of the suggested approach compared to other state-of-the-art methods in a real dataset, the relevance of each item to its corresponding students was explored from two perspectives: the group's (as a whole) and the individual student's (within the group). Results indicate a statistically significant improvement in accuracy of predicted group and individual satisfaction, as well as in the goodness of the ranked list of recommendations.**

***Keywords—collaborative learning; group recommender system; game theory; non-cooperative games; Nash Equilibrium***

## I. INTRODUCTION

In collaborative learning contexts, it is common practice to recommend resources to groups of students instead of recommending to individuals. However, recommending those educational resources to groups of learners that will optimally satisfy all group members, both as individuals and as a group, is a non-trivial task [1], [2]. The reason behind this claim is that learners in a group may not be fulfilled by the same items, yet wish to meet their own learning goals, making it difficult for them to achieve a consensus. 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 reach to an agreement of what is considered as useful, helpful, and satisfactory. The concept of "social influence" [3] explains how group members influence each other in behavioral, cognitive, and affective ways, focusing on two key processes: the emotional contagion and the conformity. Emotional contagion synopsizes how the others' emotional responses affect the individual [4], whereas conformity is the adjustment of one's opinion towards the majority [5].

The researchers in the educational group recommender systems domain, adopted the semantics originating from Social Choice Theory [6], [7], targeting mostly to model the conformity process. Although still rather sparse, the prevalent approaches include: (a) constructing groups with high member similarity and recommending resources to these groups, selecting them from a merged list of recommendations, generated for each group member separately (e.g., [8], [9]) (b) recruiting an aggregation technique for merging individual preferences in a pseudo group profile prior to generating the recommendation (e.g., [10], [11]), or (c) evaluating aggregation methods and applying classification on meta-data, including the prior evaluation results and some of the learners' characteristics [12]. In these approaches, reaching to a consensus between group members is achieved by aggregating their personalities, interests, and learning styles.

However, these methods share four types of drawbacks: (a) related to the group formation, (b) related to the aggregation strategies, (c) related to the number of recommended resources, and (d) related to the individual members' conformity degree. Regarding the first type, homogeneous groups is an unwanted restriction, since homogeneity in group formation is not always possible to be achieved. Moreover, heterogeneity in groups is considered as more beneficial for learners in collaborative learning contexts [13], [14]. Furthermore, the existing methods fail to achieve high quality performance and goodness of recommendation for majority of students in heterogeneous groups. Regarding the second type, not all aggregation strategies work efficiently in all cases, 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 type, existing methods recommend only one item per time, though it is very likely that students would possibly like to access multiple learning resources. In this case, they would be more pleased with a sequence of suggested items. Finally, regarding the fourth type, the focus of recommendation is on the overall group satisfaction, bypassing the relevance of the recommended items to the individual corresponding students, and how beneficial these items finally are to the learning subjects themselves.

Thus, the emerging research question is:

***RQ:*** *Can we accurately and efficiently recommend educational resources to homogeneous and heterogeneous groups of students, with respect to both the individuals' and the group's satisfaction?*

Towards addressing the abovementioned issues and answering the research question, we are inspired from [15] and argue that *Game Theory* could efficiently solve conflicts of interest between group members [16] and guide the recommendation of a sequence of educational resources. Game theory is "a study of mathematical models of conflict and cooperation between intelligent rational decision-makers (players)" [17]. The present work demonstrates a method for recommending educational resources to groups of students based on non-cooperative games. *Non-cooperative* is a technical term and not an assessment of the degree of cooperation among players in the game, i.e., a non-cooperative game can model cooperation, focusing on predicting individual players' choices (actions) and payoffs, but the players make self-enforced decisions independently [16].

The problem we address is how to optimally recommend educational resources to homogeneous and heterogeneous groups of students with respect to each individual member's satisfaction from the recommendation. More precisely, in our approach, the group members (students) are the players, the educational resources (items) comprise the set of possible actions, and performing a rational, self-enforced selection of items, i.e., that will maximize each group member's satisfaction (payoff), is a problem of finding the Nash Equilibrium (NE). In this state, if the other students will not modify their own actions, the student who has the option of moving away should have no incentive to unilaterally do so (the payoff doesn't improve). In case in this state the difference between the most satisfied and the least satisfied student is minimum, the solution is optimal for the group as a whole.

The remainder of the paper is organized as follows. Section II formalizes the problem of recommendation of educational resources to groups of students as a non-cooperative game. Section III presents the experimental methodology for the evaluation of our approach, and Section IV demonstrates the empirical results. Section V elaborates on our findings and contributions, and it concludes the paper.

## II. Problem Formulation as a Non-Cooperative Game

### A. Problem definition

Consider a set of learners $L$ and a set of educational resources $R$ (items). We examine the case of having learners who collaboratively solve problems in groups (at least two members). Let $G$ be a set of all groups that may be formed by $L$. If $g \in G$, then $|g|=k$, the number $k$ of group members in group $g$, with $k \geq 2$. The group members are learners with *potentially* conflicting needs and expectations. The goal is to recommend to each group those items that will be beneficial to the group as a whole and that are expected to maximize each individual member's perceived satisfaction of the items as well.

For each learner $i$ and each item $j$, the learner's perceived satisfaction $s_{ij}$ can be estimated from the learner's self-enforced evaluation of how much the particular item corresponds to the learner's expectations. Learners' perceived satisfaction from each item is measured via questionnaire in a 5-point Likert-like scale (adopted from [18], [19] – see section III); if a learner has not yet evaluated an item, then $s_{ij} = 0$. In addition, $\hat{s}_{ij}$ is the learner's $i$ *predicted* perceived satisfaction of item $j$; $\hat{s}_{ij}$ is a *decision criterion* for selecting an item $j$ given the predicted satisfaction a learner $i$ will perceive from it, and is calculated with Matrix Factorization [20] (briefly explained in next sub-section). The items to be recommended to each group should not have been previously seen or evaluated by any of the group members. We model the group recommendation problem as a non-cooperative game, i.e., a triad $(k, Q, f)$ where:

- The $k$ learners (group members) are the *players*.
- The set of unrated items $Q = \{q_z\} = \bigcap_i \{j \mid \hat{s}_{ij} = 0\}$, $Q \subseteq R$, are the available *actions*; $x = (q_1, q_2, ..., q_\mu) \in Q$ is a *strategy profile*.
- The *payoff* function for a learner $i$ and a strategy profile $x$, $f_i(x) = \frac{\sum_z \hat{s}_{iz}}{|q|}$, where $|q|$ is the total number of items in the strategy, calculates the predicted satisfaction for learner $i$ in the group, resulting from the actions by all group members – including himself – as the average individual predicted satisfaction from all items in the strategy.

The items that will be recommended to the group of learners are those in the Nash Equilibrium (NE) (single item or sequence of items). A strategy profile $x^* \in Q$ is a NE if: $\forall i, x_i \in Q : f_i(x_i^*, x^*_{-i}) \geq f_i(x_i, x^*_{-i})$, where $x_i$ is a strategy profile for learner $i$ and $x_{-i}$ is a strategy profile of all learners except for learner $i$. In other words, considering that the other learners will not modify their own strategy, the learner 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 satisfaction 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, 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 an optimum solution for the group.

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

**Algorithm 1**: Finding the Nash Equilibrium and selecting the best-response strategy

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

1. **for** i=1; i ≤ k; i++ **do** //for all students
2. assign $x_i \in Q$, $x_{-i} \in Q$ //initialize profile strategy
3. **repeat**
4. **repeat**
5. **for** i=1; i ≤ k; i++ **do** //for all students
6. assign $x_i^* \in Q$, $x_{-i}^* \in Q$ //assign another strategy
7. compute $f_i(x_i^*, x^*_{-i})$ //compute the payoff
8. **until** $f_i(x_i^*, x^*_{-i}) \geq f_i(x_i, x^*_{-i})$ //no student has incentive to //change the strategy
9. compute $d_i$ // difference max-min in NE
10. **until** min $d_i$
11. **return** $x^* \in Q$ //the NE with minimum $d_i$

Finally, a group consensus function $S(g,Q)$ computes the average satisfaction from each item in the recommended strategy $Q$ (i.e., NE) for the group $g$: $S(g,Q)=\frac{\sum_{x\in Q} f_i(x)}{|g|}$, where $f_i(x)$ is the payoff for each member $i$, and $x$ are the items in $Q$.

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

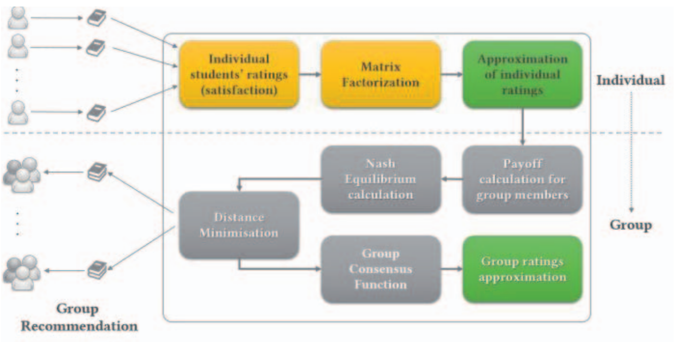



Fig. 1. Architecture of non-cooperative game-theoretic group recommender system for educational resources.

## B. *Matrix Factorization for satisfaction approximation*

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

As stated in the previous sub-section, the predicted satisfaction $\hat{s}_{ij}$ for student $i$ from item $j$ is computed with the Matrix Factorization technique [20]. The basic idea is to view the student-item satisfaction as a sparse matrix, for which we wish to predict the values of its empty cells, such that the values would be consistent with the existing satisfactions in the matrix. This is achieved by computing a low-rank approximation of the satisfaction matrix. As notational convention, bold small letters denote vectors, and bold capital letters denote matrices.

Let **S** be the matrix of size |L| x |R| that contains the satisfaction 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) satisfaction from an item $r_j$ for student $l_i$, the inner product of the corresponding factor vectors is computed: $\hat{s}_{ij}=\boldsymbol{l}_i^T\boldsymbol{r}_j$. The resulting dot product captures the student's $l_i$ overall satisfaction 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 satisfactions without over-fitting. The simplest approach to learn the factor vectors is to minimize the regularized squared error on the set of known satisfactions:

$\min \sum_{(l,r)\in K} (s_{ij}-\hat{s}_{ij})^2+\lambda(\|l_i\|^2+\|r_j\|^2)$, where K is the set of ($l_i$, $r_j$) pairs for which $s_{ij}$ 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 [21] can be applied.

## C. *k-Means clustering for group formation*

A central topic in group recommender systems is the partition of the users into a number of groups, i.e., the group formation problem. The existing methods in educational group recommender systems promote shaping homogeneous groups of students (e.g., [8], [9]). However, having heterogeneous groups of students is considered as more beneficial in collaborative learning contexts, with respect to students' overall learning gain [13], [14]. Thus, both homogeneous as well as heterogeneous groups should be considered. Furthermore, since the groups of students are not already known, supervised classification (e.g., *clustering*) techniques are appropriate for solving the problem.

The most popular clustering algorithm used in recommender systems is the k-means, mostly due to the simplicity and the efficiency that the algorithm can offer [22]. The basic idea is to group the students based on the individual ratings available (the matrix containing the students' satisfaction from the items they have already seen and rated), in such a way that students with similar ratings for the same items are in the same group. In this way, it is quite simple to end-up with homogeneous groups.

In addition, for the formation of heterogeneous groups, it is important that students have different values of the attributes considered (i.e., the ratings on the same items). A simple way to achieve this is the following: after classifying students in homogeneous groups in the previous step, students can be randomly selected from different clusters, and re-grouped in dyads and (or) triads.

# III. Experimental Evaluation

## A. *Participants and experimental setup*

The game-theoretic group-recommendation method was evaluated on a realistic setting with data from a collaborative activity with 105 students (59 girls [56.2%] and 46 boys [43.8%], aged 16 years old) from a European High School, in September 2017. The activity was about collaboratively writing simple functions in the Python programming language, using the knowledge gained during the process (as well as the lectures during the course in the classroom), and it was conducted in three phases.

During the first phase, one hundred fifty five (155) educational resources (e.g., solved exercises and worked examples), designed to motivate students and increase their interest in Python, were randomly assigned to the individuals. All students had to rate at least 3, but not more than 5 items, according to their own perceived satisfaction of each item, within 2 days. For the rating of the items, the students had to assess their own perceived usefulness of each item (adopted from [18]) and their own perceived clarity of each item (adopted from [19]), in a 5-point Likert-like scale. The average score per student was considered as the student's perceived satisfaction from the corresponding item. The resulting dataset consisted of $|S^L|=605$ student-item ratings.

Before the initiation of the second phase, the students were arranged into four general, equivalent groups: one treatment (*E* – 27 students) and three control groups (*C1, C2, C3* – 26, 25, 27 students respectively). Each of these general groups was further partitioned in homogeneous and heterogeneous sub-groups. For the sub-group formation, k-means clustering was applied (described in section II.C); initially, students with similar ratings on the same items were selected to be in the same groups, and then for the formation of the heterogeneous groups, students from the different clusters were organized together. $|G|=9$ sub-groups, including 4 or 5 homogeneous and 4 or 5 heterogeneous (i.e., 36 sub-groups in total), with $|g|$ varying from 2 to 3 students per sub-group, were finally formed for each one of the general groups.

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

TABLE I. DESCRIPTION OF GROUPS IN SECOND PHASE

| GroupID | No. of students | Homogeneous | Heterogeneous | Rec. Method |
|---|---|---|---|---|
| *E* | 27 | 4 | 5 | GT |
| *C1* | 26 | 5 | 4 | AVG |
| *C2* | 25 | 4 | 5 | LM |
| *C3* | 27 | 5 | 4 | MP |

After studying the recommended items for two days, all group members had to rate them both individually, and as a team, using the same assessment method (questionnaires) as in the previous phase. Every second day, the matrixes containing the real individual ratings and the actual group ratings, were updated. At the end of the second week, the learner-item ratings were $|S^L|=1133$, and the group-item ratings were $|S^G|=196$. Throughout the experimental process, the items recommended to each sub-group should not have been previously seen and rated by any of the sub-group members.

Finally, the third phase of the activity was about collaboratively writing simple functions in Python, using their knowledge gained during the previous two phases, as well as the available educational resources, within one week time.

The activity cycle is illustrated and synopsized in Fig. 2.

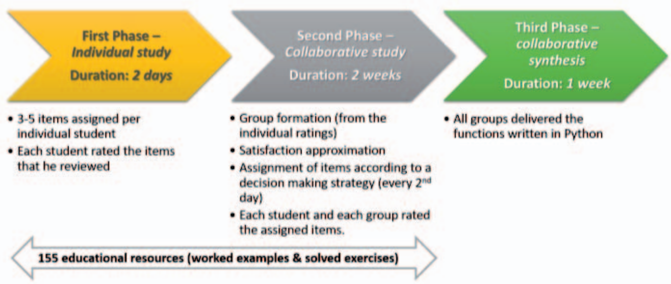



Fig. 2. The experimental activity process.

### B. *Methods and Evaluation metrics*

*1) Group decision strategies.* As stated in the previous sub-section, for each one of the control groups (i.e., C1, C2, C3), the expected group satisfaction from an item was provided by a different group decision (aggregation) method, formulating how the corresponding sub-groups of students reach to a consensus and come up with a decision about that particular item. More precisely, the group satisfaction 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. Let *k* be the number of students in a group, $s_{ij}$ the satisfaction of student *i* from item *j*, then the group satisfaction equals the average satisfaction ratings across the group members: $S(k,j)=\frac{\sum_{i\in g} s_{ij}}{k}$. 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 – Least-Misery (LM):*** A borderline approach that targets to please the least happy member of the group, resulting the group to behave under a least-misery principle. In this case, the group satisfaction equals the minimum satisfaction among all group members: $S(k,j)=\min_{i\in g} s_{ij}$. In other words, *LM* considers the rating of each item, and then 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 satisfied as its least satisfied member.

***C3 – Most-Pleasure (MP):*** Another borderline group decision strategy satisfying the highest rating within the group. The satisfaction a group of *k* students gets from an item *j* equals the maximum satisfaction within the group: $S(k,j)=\max_{i\in g} s_{ij}$. Similarly to LM, MP takes under consideration the ratings of each item and next recommends the item with the maximum satisfaction between all group members. Thus, a group is as satisfied as its most satisfied member.

All solutions were implemented in MATLAB. Furthermore, the Gambit tool [24] 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 satisfaction 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 [12]. In our study, we explore the precision of our prediction with respect to satisfaction from the recommended items, as it is actually rated by *each* student, and by a given *group* of students. RMSE is computed as: $RMSE = \sqrt{\frac{\sum_{j=1}^{n}(s_{kj} - \hat{s}_{kj})^2}{n}}$, where $n$ is the number of items rated. Lower values indicate better predictions, and consequently, better decision strategy.

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 measures from Information Retrieval, specifically crafted for ranking: the *Normalized Discounted Cumulative Gain (nDCG)* which assumes multiple levels of relevance [14]. The main distinction of DCG from other measures is the ability to address non-normal rank distribution by assigning a higher cost to high-rank elements. This emphasizes the high-rank element identification.

In simple terms, *Discounted Cumulative Gain (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. 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: $DCG_k@N = s_{kj_1} + \sum_{i=2}^{N}\frac{s_{kj_i}}{log(i+1)}$. However, comparing DCGs between groups of students is not valid. As such, normalized DCG (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: $nDCG_k@N = \frac{DCG_k@N}{IDCG_k@N}$, where IDCG is the maximum possible DCG, and $nDCG_k@N$ 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 N=3 and N=5. We compared the effectiveness of both the *group* and *individual* recommendations when varying the aggregation method and the group size (equal to 2 and 3).

Finally, we measure the diversity of recommendation lists between different groups, by employing the *Hamming Distance (HD)* [23] 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 $HD(g,g^*) = 1 - \frac{Q_{g,g^*}}{|z|}$, where and $z$ 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. Generally speaking, a highly personalized recommendation list should have higher HD to other lists.

## IV. RESULTS

Tables II and III demonstrate the results for the evaluation measures (average values) for all decision support strategies compared in this study, i.e., the currently proposed game-theoretic method (*GT*) applied on the treatment group, and the Average (*AVG*), Least-Misery (*LM*), and Most-Pleasure (*MP*) methods applied on each one of the control groups, for homogeneous (high inner sub-group similarity), as well as heterogeneous (low inner sub-group similarity) synthesis of the sub-groups respectively. The sub-groups sizes was firm, varying from two to three students, as explained in section III.A.

TABLE II. PREDICTION ACCURACY AND GOODNESS OF RANKED LIST OF RECOMMENDATIONS FOR HOMOGENEOUS GROUPS

| | *RMSE* | *nDCG@3* | *nDCG@5* | *HD* |
|---|---|---|---|---|
| ***GT*** | **0.348** | 0.967 | **0.969** | **0.862** |
| ***AVG*** | 0.352 | **0.968** | 0.968 | 0.774 |
| ***LM*** | 0.373 | 0.889 | 0.887 | 0.625 |
| ***MP*** | 0.412 | 0.884 | 0.879 | 0.618 |

TABLE III. PREDICTION ACCURACY AND GOODNESS OF RANKED LIST OF RECOMMENDATIONS FOR HETEROGENEOUS GROUPS

| | *RMSE* | *nDCG@3* | *nDCG@5* | *HD* |
|---|---|---|---|---|
| ***GT*** | **0.432** | **0.934** | **0.923** | **0.796** |
| ***AVG*** | 0.647 | 0.852 | 0.851 | 0.714 |
| ***LM*** | 0.722 | 0.737 | 0.726 | 0.612 |
| ***MP*** | 1.206 | 0.689 | 0.624 | 0.534 |

Similarly, tables IV and V synopsize the results for the evaluation metrics (average values) for the individual students in each of the sub-group categories, respectively.

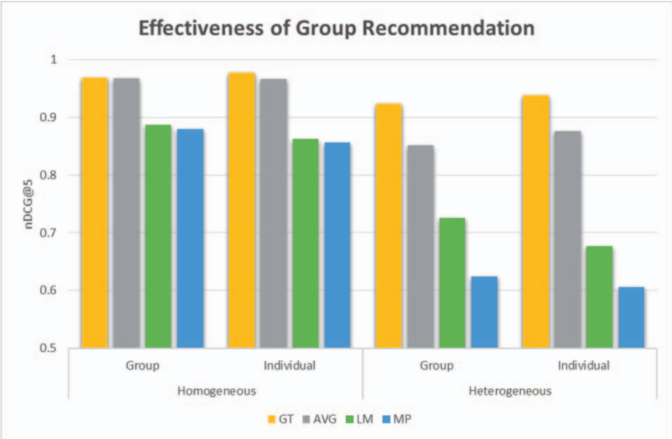
TABLE IV. PREDICTION ACCURACY AND GOODNESS OF RANKED LIST OF RECOMMENDATIONS FOR INDIVIDUAL STUDENTS IN HOMOGENEOUS GROUPS

| | *RMSE* | *nDCG@3* | *nDCG@5* | *HD* |
|---|---|---|---|---|
| ***GT*** | **0.346** | **0.972** | **0.977** | **0.862** |
| ***AVG*** | 0.428 | 0.969 | 0.966 | 0.774 |
| ***LM*** | 0.483 | 0.875 | 0.863 | 0.625 |
| ***MP*** | 0.506 | 0.867 | 0.856 | 0.618 |

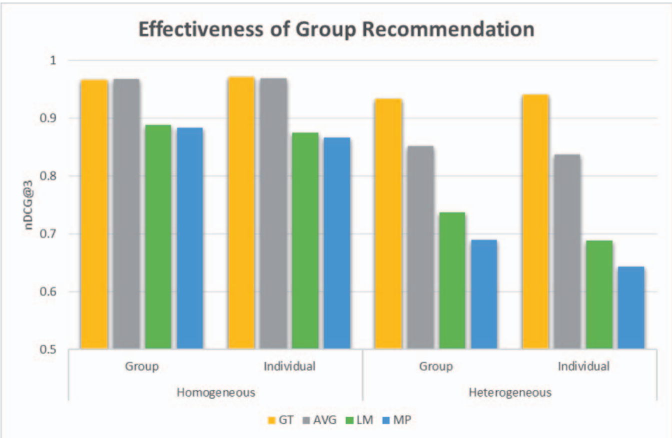
TABLE V. PREDICTION ACCURACY AND GOODNESS OF RANKED LIST OF RECOMMENDATIONS FOR INDIVIDUAL STUDENTS IN HETEROGENEOUS GROUPS

| | *RMSE* | *nDCG@3* | *nDCG@5* | *HD* |
|---|---|---|---|---|
| ***GT*** | **0.394** | **0.941** | **0.938** | **0.796** |
| ***AVG*** | 0.686 | 0.837 | 0.876 | 0.714 |
| ***LM*** | 0.830 | 0.688 | 0.677 | 0.612 |
| ***MP*** | 1.454 | 0.643 | 0.606 | 0.534 |

According to these results, all decision support methods achieve low approximation error in prediction of satisfaction ratings for the homogeneous students' sub-groups. On the contrary, for heterogeneous sub-groups, accuracy is high for the GT and AVG methods, but the prediction error significantly increases when the aggregation strategy is LM or MP. Furthermore, the group recommendations effectiveness tends to decrease only for the heterogeneous sub-groups. Figure 3 illustrates the goodness of the ranked list of recommended items delivered to the sub-groups of students (a) when the top ranked items are 5 (nDCG@5) and (b) when the top ranked items are 3 (nDCG@3), according to the inner similarity of the sub-groups, and by considering the decision support strategy.




(a) 5-top ranked recommendation items




(b) 3-top ranked recommendation items

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

In addition, the diversity in recommendations, as reflected in the HD values, indicates that the recommendations to all groups (both the homogeneous and the heterogeneous) make full 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 heterogeneous groups.

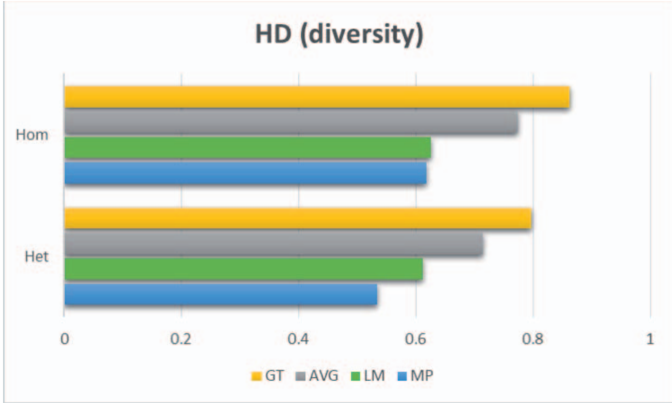



Fig. 4. Diversity of group recommendations with respect to the aggregation strategy and the group inner similarity.

Furthermore, in order to understand *when* the group recommendations are better or worse (ranked) for each individual within the sub-groups, we measured the difference between the effectiveness of the individual and the group recommendations' lists. This difference is indicative of the individuals' degree of conformity, regarding the adjustment of their satisfaction from the recommendation with respect to the satisfaction of the group they are members of. A positive difference means that the group recommendations are better ranked than the individual recommendations. Fig. 5 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. 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. In this figure, the green trendline corresponds to the GT method, whereas the red trendline corresponds to the AVG method, respectively.

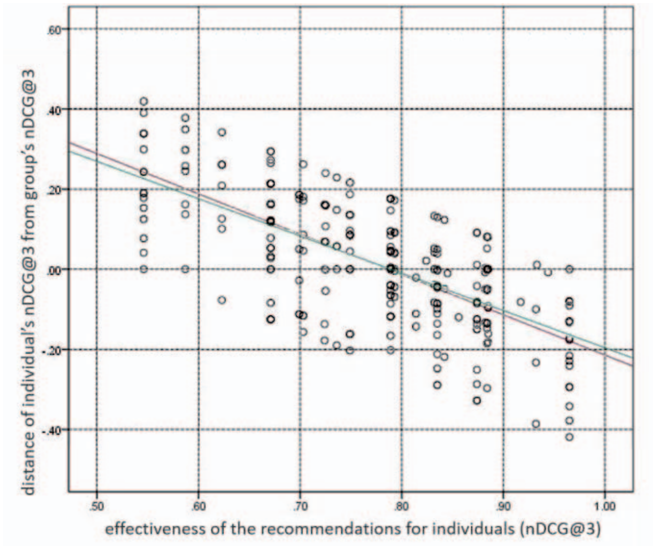



Fig. 5. Distance of group satisfaction with respect to individual satisfaction.

## V. Discussion And Conclusions

Recommending educational resources (items) to groups of students, targeting at optimizing all students' satisfaction, both individually and as a group, is a complicated task. The core issue is to determine how a group of students reaches a consensus about the rating for each item in a way that reflects the interests and satisfaction of each group member. This study focuses on solving conflict of interest among students and recommending educational resources to groups in online collaborative learning contexts, and follows a non-cooperative game-theoretic perspective.

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). However, the best collective result does not always come from each individuals following their own interest, but rather from reaching the group's consensus.

In this paper, we developed a mathematical formulation (i.e., algorithm1) for group-recommendation of educational resources as a non-cooperative game, as well as an architecture for building such group recommender systems (i.e., Fig. 1). The proposed solution models the recommendation strategy as a problem of finding the Nash Equilibrium, i.e., a state in which no student can be benefited more in terms of further improving their own satisfaction by unilaterally deviating from the N.E. We also introduced the influence of the conflict among students during collaboratively making decisions, as a factor of students' degree of conformity; we explicitly compared the difference in individual evaluation of satisfaction from the recommendation, from the group's perception.

An empirical study with a realistic dataset was conducted for the evaluation of the suggested approach. The goal was to compare the performance 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 satisfaction from the recommendation. The following novel facts and important observations have risen.

Firstly, all decision support methods achieve low approximation error in prediction of satisfaction ratings for the homogeneous students' sub-groups. On the contrary, for heterogeneous sub-groups, accuracy is high for the GT method, but the prediction error significantly increases when the aggregation strategy is AVG, LM or MP. More precisely, from tables II and III, it becomes apparent that the proposed game-theoretic strategy minimizes the prediction error of the sub-group satisfaction ratings, as, by far, it scores the lowest RMSE values for all categories of inner sub-group similarity. Especially for the highly heterogeneous sub-groups, the other aggregation methods combine potentially conflicting rankings that could create a group recommendation which might not be satisfactory for the group members. In this case, the GT decision strategy resolves sufficiently the conflict of interest and delivers the most appropriate items to the students.

Similar are the findings from the individual level of analysis: as seen from tables IV and V, the predicted individual satisfaction from the recommended items is accurate for all methods when the learner is a member of a homogeneous group, but is more accurate with the GT method, when the student participates in heterogeneous groups.

Secondly, we also observe that our 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 more effective recommendations). 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 sub-group 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 sub-group homogeneity, whereas the respective values for the other methods decrease as the inner group similarity decreases.

Thirdly, another finding concerns the diversity of the recommended lists of items. The values of the Hamming Distance (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.

Fourthly, in order to understand how much the individuals adjusted their personal evaluation of satisfaction from the recommended items compared to the group's they belong to, one can observe that when the employed method is the GT, the individual recommendations are better ranked than the group recommendations. In other words, the individuals within the groups don't have to highly adjust their personal consideration about their satisfaction from the recommended items. The measure employed, i.e., the distance between the individual nDCG and the respective group's nDCG, could be further explored as a measure of the individual's degree of conformity.

However, there are some limitations. Firstly, the samples of the 155 educational resources and 105 students considered in the evaluation process are small; bigger datasets should be analyzed. Secondly, 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. Lastly, 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.

Furthermore, a number of challenges for future work has emerged. For example, more sophisticated measures of satisfaction could be applied (e.g., incorporating the students' affective states, perceived enjoyment, challenge). The learning analytics research could contribute towards this direction. Yet, another challenging issue is focusing on the transparency of the group recommendation: showing each individual's payoff and eventually, how satisfied 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.

To conclude, the contribution of this study is that the proposed solution demonstrates a socially and individually optimum group recommendation method, beyond aggregation of individual profiles or merging of individual recommendation approaches, 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 *Proceedings of the Fourth ACM Conference on Recommender Systems*, 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: Springer US, 2011, pp. 677–702.

[3] 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.

[4] E. Hatfield, J. T. Cacioppo, and R. L. Rapson, "Emotional Contagion," *Curr. Dir. Psychol. Sci.*, vol. 2, no. 3, pp. 96–100, 1993.

[5] S. E. Asch, "Effects of group pressure upon the modification and distortion of judgments.," in *Groups, leadership and men; research in human relations.*, Oxford, England: Carnegie Press, 1951, pp. 177–190.

[6] D. M. Pennock, E. Horvitz, and C. L. Giles, "Social Choice Theory and Recommender Systems: Analysis of the Axiomatic Foundations of Collaborative Filtering," in *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, 2000, pp. 729–734.

[7] P. C. Fishburn, *The Theory of Social Choice*. Princeton University Press, 1973.

[8] 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.

[9] D. Yanhui, W. Dequan, Z. Yongxin, and L. Lin, "A Group Recommender System for Online Course Study," in *2015 7th International Conference on Information Technology in Medicine and Education (ITME)*, 2015, pp. 318–320.

[10] 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.

[11] P. Rodríguez, M. Giraldo, V. Tabares, N. Duque, and D. Ovalle, "Recommendation System of Educational Resources for a Student Group," in *Highlights of Practical Applications of Scalable Multi-Agent Systems. The PAAMS Collection: International Workshops of PAAMS 2016, Sevilla, Spain, June 1-3, 2016. Proceedings*, J. Bajo, M. J. Escalona, S. Giroux, P. Hoffa-D\kabrowska, V. Julián, P. Novais, N. Sánchez-Pi, R. Unland, and R. Azambuja-Silveira, Eds. Cham: Springer International Publishing, 2016, pp. 419–427.

[12] 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.

[13] S. Graf and R. Bekele, "Forming Heterogeneous Groups for Intelligent Collaborative Learning Systems with Ant Colony Optimization," in *Intelligent Tutoring Systems: 8th International Conference, ITS 2006, Jhongli, Taiwan, June 26-30, 2006. Proceedings*, M. Ikeda, K. D. Ashley, and T.-W. Chan, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 217–226.

[14] R. E. Slavin, "Developmental and Motivational Perspectives on Cooperative Learning: A Reconciliation," *Child Dev.*, vol. 58, no. 5, pp. 1161–1167, 1987.

[15] L. A. M. C. Carvalho and H. T. Macedo, "Users' Satisfaction in Recommendation Systems for Groups: An Approach Based on Noncooperative Games," in *Proceedings of the 22Nd International Conference on World Wide Web*, 2013, pp. 951–958.

[16] J. Nash, "Non-Cooperative Games," *Ann. Math.*, vol. 54, no. 2, pp. 286–295, 1951.

[17] R. B. Myerson, "Negotiation in Games: A Theoretical Overview," Jul. 1985.

[18] F. D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Q.*, vol. 13, no. 3, pp. 319–340, Sep. 1989.

[19] V. Terzis and A. A. Economides, "The acceptance and use of computer based assessment," *Comput. Educ.*, vol. 56, no. 4, pp. 1032–1044, 2011.

[20] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based Collaborative Filtering Recommendation Algorithms," in *Proceedings of the 10th International Conference on World Wide Web*, 2001, pp. 285–295.

[21] L. Bottou, "Large-Scale Machine Learning with Stochastic Gradient Descent," in *Proceedings of COMPSTAT'2010: 19th International Conference on Computational StatisticsParis France, August 22-27, 2010 Keynote, Invited and Contributed Papers*, Y. Lechevallier and G. Saporta, Eds. Heidelberg: Physica-Verlag HD, 2010, pp. 177–186.

[22] X. Amatriain, A. Jaimes, N. Oliver, and J. M. Pujol, "Data Mining Methods for Recommender Systems," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Boston, MA: Springer US, 2011, pp. 39–71.

[23] T. Zhou, L.-L. Jiang, R.-Q. Su, and Y.-C. Zhang, "Effect of initial configuration on network-based recommendation," *EPL (Europhysics Lett.*, vol. 81, p. 58004, Mar. 2008.

[24] R. D. McKelvey, A. M. McLennan, and Th. L. Turocy. Gambit: Software Tools for Game Theory, Version 16.0.1. http://www.gambit-project.org., 2016.