# International Journal of Advanced Trends in Computer Science and Engineering

# Towards an Improvement of Teaching by Preventing Courses Based on Learning Communities

**Salhi Intissar[1], Qbadou Mohammed[2], Mansouri Khalifa[3]**
Laboratory Signals, Distributed Systems, and Artificial Intelligence ENSETM University Hassan II.
Casablanca Morocco,
salhi9477@gmail.com[1], qbmedn7@gmail.com[2], khmansouri@hotmail.com[3]

## ABSTRACT

In each semester and after the first lesson sessions, teachers readjust their lessons and the way they present them to suit the learning profiles of learners (the understanding, motivation, and identity profile), this constitutes a very difficult task. In the personalized learning, the question does not arise. This is why this mode of learning is so effective. With the continuous improvement of learning systems and the very large amount of information that we can take advantage of. Previous research in data mining has identified several models that can be applied to predict course selection based on data residing in institutional information systems. However, these models aim to analyze only the historical data of students such as course registrations, course load, and academic results to determine the probability of success and cannot act during the course period. In this article, we propose an approach to extract student preferences from the behavior of the learning community to which they belong. We applied a community detection algorithm to a set of students based on their activity that we validated by analyzing scores and used to determine the preferences of every learning community to target lessons and to improve decision making.

**Key words:** Learning profiles, personalized learning, Learner activity, institutional information systems, student learning communities.

## 1. INTRODUCTION

It is a common fact that technological tendencies bring about continuous trade in all sectors of current society. Education itself cannot remain passive and indifferent; all typical educational techniques are revised and reassessed and new ones are introduced. E-learning systems strive to meet the current academic needs of learners.

In learning management systems (LMS), the personalization of learning consists of adapting educational resources to a learner or a group of learners with the same characteristic. In this case[1], we talk about the personalization of educational resources.

The personalization of pedagogical resources to learners can be done in the context of individualized learning where the pedagogical resources are adapted to the goals and needs of each learner, according to their characteristics, or personalized learning [2][3] in which the learner works with his teacher to set short and long-term goals and the resources which seem relevant to him after having carried out a reflexive activity on himself and his learning.

To meet this goal, various e-learning systems have been developed in recent years, however, most of them form old-fashioned static applications, lacking features such as behavior analysis during classes, and detecting learning profile. Since the majority of e-learning systems [4][5][6][7][8] consider the learner as an entity accompanied by a predefined static set of interests and options, without paying attention to their needs. The lack of automatic learner profile assessment and the lack of the provisioning of an appropriate learning resource is a common phenomenon in these systems. Besides, none of these systems performs a statistical analysis of profiling to improve the performance of learners.

Students now complete the usual course structure with online material, they can view a course, participate in an online discussion forum, or answer questionnaires and do their homework online. This, in turn, gives a multitude of new behavioral data relating to the profiles of learners. These data evolve gradually as the learning progresses. And by exploiting the traces left by the learner on the networks, we can group the students into communities using standard community detection algorithms to create qualitative and practical software systems that will allow instructors to continually improve their teaching approaches.

Student learning communities are defined as any Purposely designed curricular program where the same group of students takes a common set of courses together, which are connected in the same meaningful way or share a common curricular experience [9]. In short, a Student Learning Community is a curricular-based learning-centered peer-to-peer social network that extends beyond the classroom.

In this article, we propose a personalization of learning and an adaptation of teaching resources to learner's needs based on the results of a recent study carried out by our team which consists of detecting students learning communities according to centrality. In our approach, the key point is to determine the interests of the learners through a pre-evaluation process, to follow their progress and their behavior during the lesson sessions, and to filter the educational material proposed according to a statistical study on the learning communities that we detect using community's detection techniques. This allows us to provide teachers with a more effective decision support tool. We begin by explaining the main elements of this study, namely the method that we designed and analyzed, the database that we used, as well as the experimental approaches used for the adaptation of resources and classes. After that, we continue with a discussion of the obtained results. And we conclude with some recommendations for future studies.

## 2. THEORETICAL BACKGROUND

### 2.1. Predicting course selection

There are many examples in the research literature illustrating the use of data mining techniques to develop course recommendation systems[10][11]. In the article[12][13][14] the authors determined the probability of success of a student in a selected course by analyzing his historical data, his course registrations, the course load, and the academic results.

In the papers[15][16][17], the authors talked about the importance of processing learning data to optimize learning opportunities for students.

In [18], the authors developed a new recommendation system based on a recurrent neural network to suggest courses to help students prepare for targeted courses of interest, personalized according to their prior knowledge and the area of proximal development.

### 2.2. Student learning communities' detection

The detection of communities in an education network consists of discovering groups of learners who maintain a distinctive profile so that they all have at least the same level of competence. Moreover, finding out who is the most influential person (leader) may be very relevant and helpful for the instructor (learning agent) as it will assist him/her in the interaction with the learners. Using this data analysis approach, instructors will be able to track their trainees, including their behavior, their overall performance, and their level of satisfaction with the lessons. It will also help them to predict the level of their trainees and their capabilities by focusing solely on each community leader because each one's behavior is similar to that of the members of the community in which he or she is the leader. Thus, making it possible to

enhance the learning experience by making well-informed and optimal recommendations and then providing a list of potential learning materials and recommendations.

Intending to quantify the importance of a node in a graph, researchers have proposed several definitions which are known as centrality measures. Identifying the central nodes within a graph represents a key issue in many areas. In the case of educational networks, identifying important nodes allows us therefore better to control the nodes which belong to the same community with minimal effort. Among these measures is the betweenness centrality.

Betweenness Centrality[19] is a measure of global centrality proposed by Freeman. The intuition of this measure is that, in a graph, a node is all the more important because it is necessary to cross it to go from one node to another. More precisely, a vertex with a strong betweenness centrality is a vertex through which a large number of geodesic paths pass the graph. In a social network, an actor with a strong betweenness is a vertex such that a large number of interactions between non-adjacent vertices depend on it [[20]. In a communication network, the betweenness centrality can be considered as the probability that information transmitted between two nodes, which passes through this intermediate node.

To calculate the betweenness centrality, we count geodetic distance $\delta_{ij}$ separating actors i and j, and look at the $\delta_{ij}(m)$ number passing through the actor m.

$$C_b(m) = \frac{2}{(n-1)(n-2)} \sum_{i \neq m} \sum_{i < j \neq m} \left(\frac{\delta_{ij}(m)}{\delta_{ij}}\right) \qquad (1)$$

More $C_b$ is higher, higher is the central summit as it is located at the junction.

## 3. OUR APPROACH

### 3.1. The detection of learning communities using betweenness centrality

In our recent papers[21], [22], we have found out that in a smart school classroom that is equipped by sensors as well as cameras, through using web-based solutions as a Learning Management System, we can build up new social networks and link up the social interactions between students from different classes via the Internet. It has allowed us understanding how this complex network is structured and to extract meaningful knowledge out of it. For this purpose, we have reported our research on the detection of students' learning communities based on their activity and have found that we can group them in communities by the content of their messages as well as by their responses structures when applying standardized Community detection algorithms

[23][24][25]. Moreover, their behaviors may be highly interrelated with their closest pairs that belong to the same community.

We used One of the most used approaches. The Girvan Newman approach uses edge-betweenness (1) to infer groups. This is the corollary of node betweenness, which is a measure of centrality, Girvan Newman extends this notion to edges. This algorithm offers quality results and intuitive reasoning by following the steps below:

- Step 1: We calculate the edge-betweenness of all edges.
- Step 2: The edges with the strongest edge-betweenness, i.e. which are most likely to be bridged between communities, are removed.
- Step 3: These two steps are repeated.

The main drawback is a high algorithmic complexity, which limits its use to small graphs.

Let G = (X, U) an n-order graph (oriented or not). The edge betweenness of a node m is defined by the formula (2):

$$E_b(m) = \sum_i \sum_{i>j} \frac{\delta_{ij}(m)}{\delta_{ij}} \tag{2}$$

where $\delta_{ij}(m)$ is the total number of geodesic paths between nodes $m_i m_j$ and that pass through the node m, and $\delta_{ij}$ is the total number of geodesic paths between nodes $m_i m_j$ and?

### 3.2. Students' Academic Performance Dataset

the Student Academic Performance Data Set (xAPI-Edu-Data) is an educational dataset that is collected from a Learning Management System known as Kalboard 360, a multiagent LMS that has been designed as a tool that facilitates learning by the use of advanced technologies. Such a system offers learners synchronized access to learning materials from any internet-connected device.

The data set is composed of 480 learner records and 16 features. Characteristics are categorized into three main classes: (1) demographic attributes like gender or nationality. (2) Academic attributes including educational level, grades, and sections. (3) behavioral attributes including class handing over the classroom, opening up resources, parent surveys, and school satisfaction.

The dataset consists of 305 males and 175 females from different backgrounds. The dataset has been collected over two semesters of schooling: during the first semester 245 student records are collected and 235 students' records were collected in the second semester.

Students were classified by their total grade/score into three numerical bands that are low level describing the interval that includes values from 0 to 69, intermediate level with values from 70 to 89, and high level with values from 90 to 100. It is on scores we will focus to reveal a correlation of students from the same community.

### 3.3. Communities discovery

We build the graph shown in Figure 1 with Newman-Girven's approach. Where each node in the social network is a student of the class. And relationships between students appear as arcs. Relationships between students can be defined in terms of higher degree similarity.

After the network has been built, we selected the central nodes that are reflected in the graph with appropriate colors. See figure 2. and for the discovery of the communities, we get a clear (net) distribution, i.e. one individual is part of a unique and unique community. Refer to figure 3.
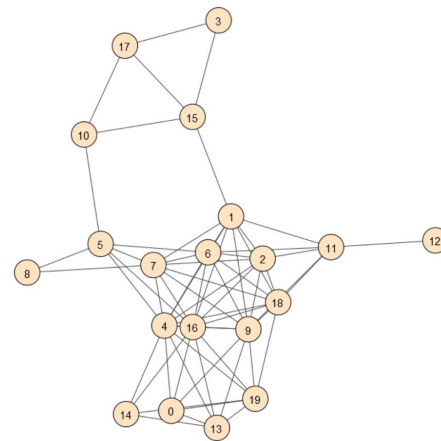


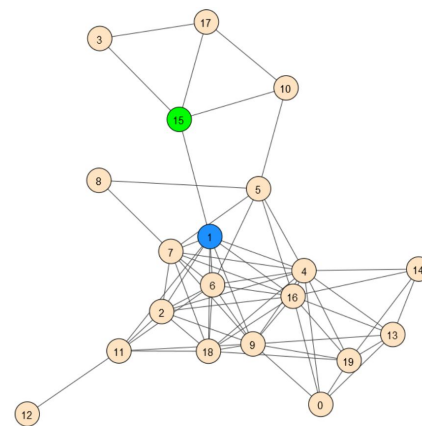**Figure 1:** Social Network built with 20 students randomly selected.



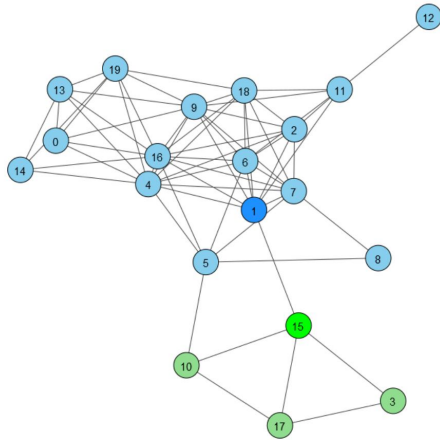**Figure 2:** Student Social Network with Highlighted central student nodes.

**Figure 3:** graphical illustration with two highlighted communities.

## 4. RESULTS AND DISCUSSION

### 4.1. the statistical analysis results to validate the approach

In all applied studies, one of the basic steps to validate the study should be the description of and exploration of the data available, before drawing up rules and predictive models. To do so, the most popular graphical multivariate statistical tools were used. The results are presented in the next sections.

*1) Scatterplot Matrix*

In statistics, a scatterplot[26] represents the data depending on a set of parameters. It shows up the degree to which two or more related variables are correlated. Observations from the scatterplot allow us therefore to identify tendencies, interdependencies, positive, negatives, directs, indirect or inverts dependencies, distributions that are more and less homogeneous, deviations from the standard deviation, outliers or subgroups that may correspond to the application of a normal distribution. And if we have multivariate data, we use the scatterplot matrix.

Looking at the paired graph for the students' first community of learning (Figure 4), it is quite apparent that the medium- and low-level learners differ in each paired graph combination, but the higher-level learners are dispersed pretty much everywhere and are blended with the medium- and the low-level ones.

But in the pair-wise illustration below for the second learning community (Figure 5), high-level students are recognizable and it is pretty clear that this learning community is defined by high-level learners.

That means, we should contribute further study on the selection of the number of communities, in such a way as to optimize community recognition in a group of students, so that it can be incorporated into an educational recommendation system.
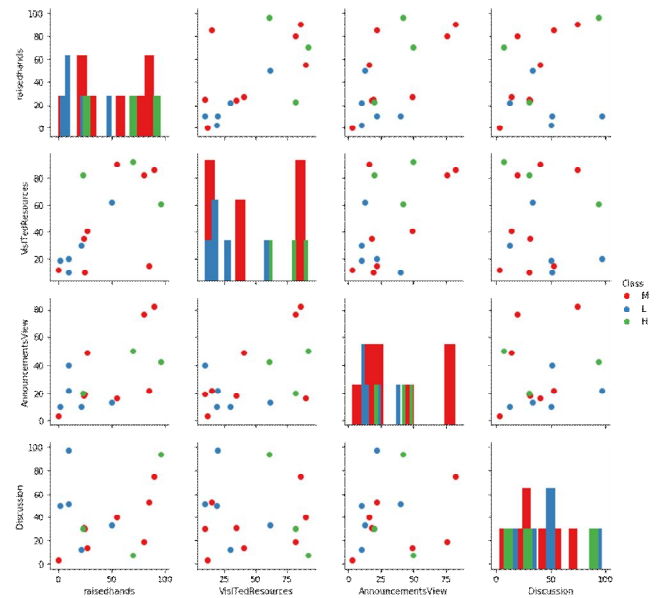


**Figure 4:** Scatterplot matrix for the axes of the attributes 2-2 for the first learning community.
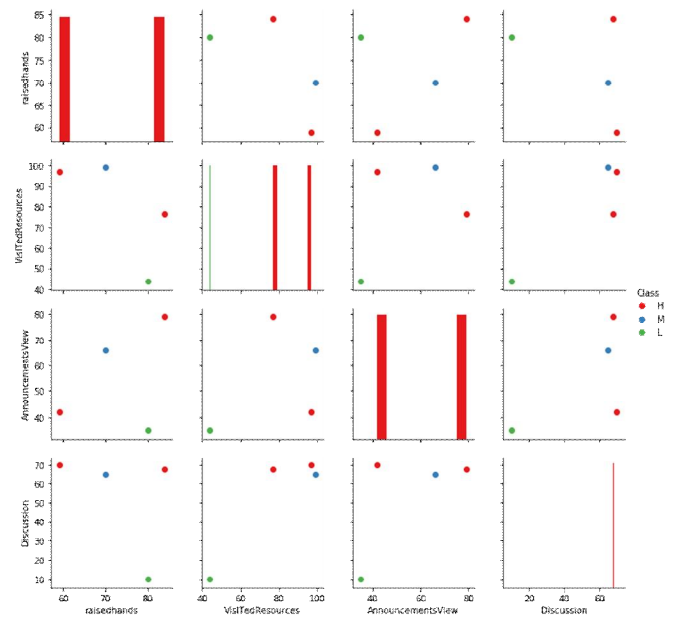


**Figure 5:** Scatterplot matrix for the axes of the attributes 2-2 for the second learning community.

*1) The boxplot*

The Boxplot graph [27] represents the most popular type of statistical chart and one of the rare and widely used statistical

chart types. Thanks to their elegance as well as to their functionality, Boxplots offer a multitude of variations and refinements.

The Boxplot consists of a small graph that represents a variety of dispersive features in a statistical sequence. This is a very quick and easy way to represent the essential profile of a quantitative statistical series and it is often used for comparison purposes. It was first invented in 1977 by John Tukey[28] but it has to be adapted according to the users.

The two graphs in figure 6 and figure 7 indicate that in the first community, the three levels are present, but that it is the medium level which reigns. And for the second community, there is only a high level.
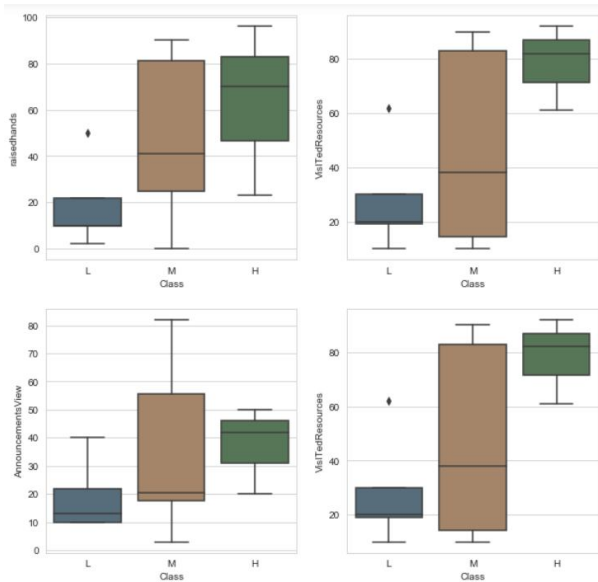


**Figure 6:** The boxplot chart for the first Student Learning Community.
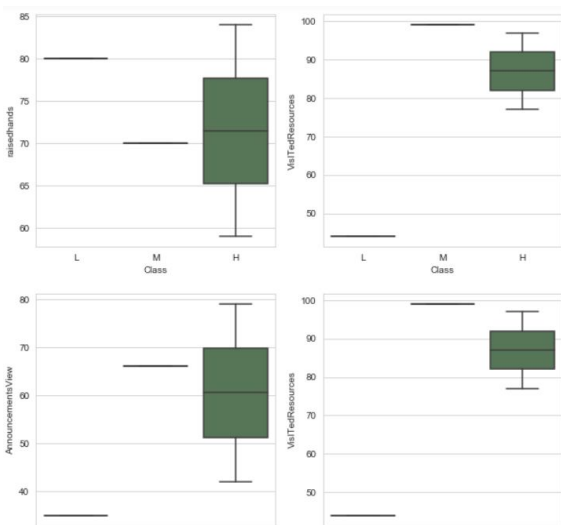


**Figure 7:** The boxplot chart for the first Student Learning Community.

## 4.2. Courses Prediction based on their activity

To predict courses for the next Semester and to recommend a program to the learners, we applied some Exploratory Data Analysis Tests.

*1) Bar Plot*

Figures 8 and 9 below illustrate a Bar Plot charts of Raise hands Average in terms of Topics, for the first and the second communities respectively.
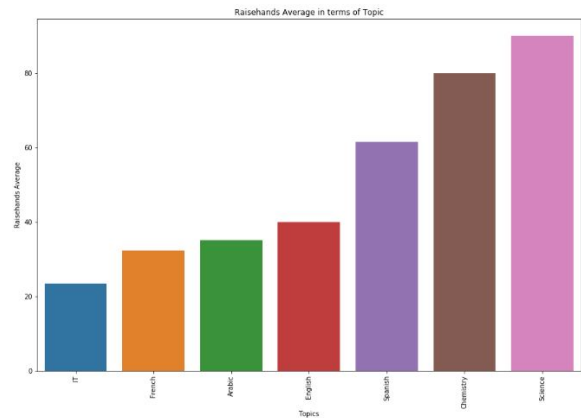


**Figure 8:** Raise hands Average in terms of Topics for the first learning community.

Figure 8 clearly shows that the learners who belong to the first community raise their hands especially in science subjects and especially in the science and chemistry course as well as the Spanish language. While the rest of the subjects are not. This allows us to deduce that the learners from the first community are more interested in science, chemistry, and the Spanish language, and perhaps they have difficulties in following and understanding the content of the rest of the courses.
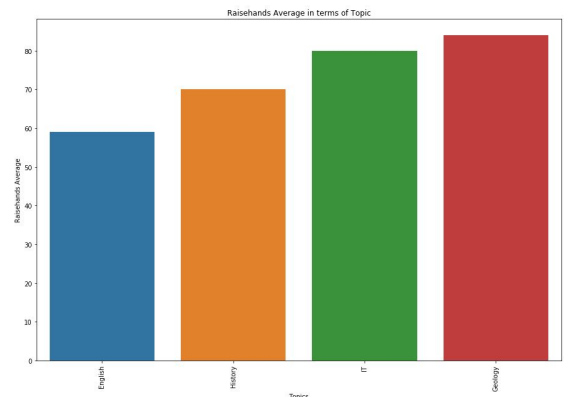


**Figure 9:** Raise hands Average in terms of Topics for the second learning community.

In Figure 9 we see that learners from the second community this time raise their hands in geology, history, and IT classes. This allows us to deduce that learners from the second community will be more interested in its courses than others. Based on the results of the two figures 8 and 9, we can conclude that for the rest of the pedagogical path of the learners we have to think about:

- Directing the students of the first community towards a purely scientific class.
- Direct the students of the first community towards a purely Geology-history path.
- Adapt courses that are not responsive to the needs of the students to help their understanding.

*2) Point Plot*

The figures 10 and 11 below illustrate the point plot charts of Raise hands Average Raised hands vs Discussion Rate.
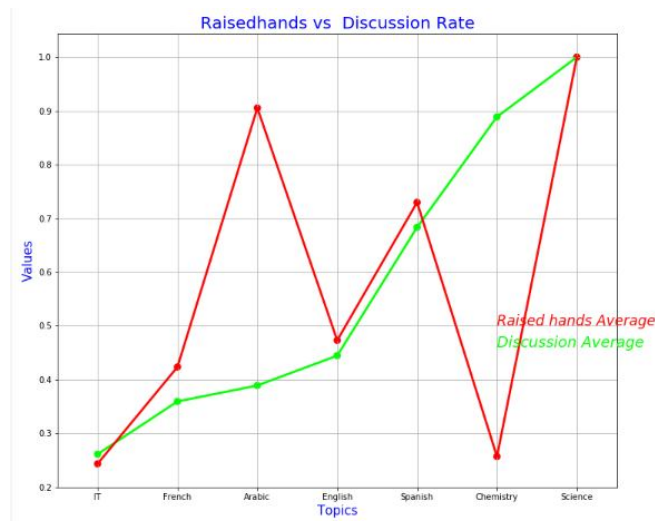


**Figure 10:** Raise hands vs Discussion rate for the first learning community.

As Figure 10 demonstrates, learners of the First Community are more likely to be more responsive during the Arabic Language course but do not express many conversations about the topics of the lesson. While for the Chemistry class they tend to talk more about the lecture in the discussion forum, but their activities during the course are very minimal.

As for Figure 11, it shows us that students from the second community tend to be more reactive and more likely to pose questions and responses during IT class.

Based on the results of Figures 10 and 11, we can say that for:

- The first community: the students are very interested in the chemistry class since they discuss it a lot, but they do

not respond during the lecture session. This can be because of the very strict attitude of the instructor or because of the complexity of the subject. However, during the Arabic language class students are very reactive and they do not have to discuss their difficulties in the forum, this implies that the teacher's teaching strategy was the right one.

- The second commonality: the students do not react much with the IT class, which is obvious because it does not belong to the same category of classes to which the students are more responsive, therefore we have to think about removing the class from the academic paths of these students.
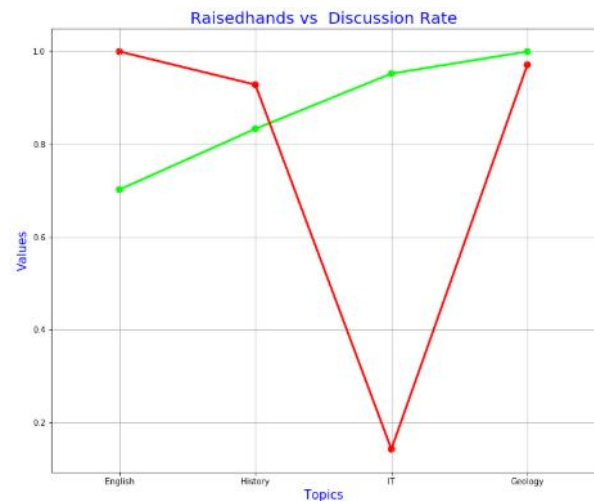


**Figure 11:** Raise hands vs Discussion rate for the second learning community.

## 5. CONCLUSION

As a result of the recent technological progress, massive quantities of data are being accumulated at a rapid pace in different sectors of human activity. These include the activity of learning. The understanding of both universal and the specific aspects of the networks related to this data has now become a real and important challenge. Understanding the community organization helps predict certain key aspects of the systems under study. For instance, using our approach, students' learning communities within the educational network may be discovered allowing teachers the recommendation of an instructional program and may be beneficial for them so they can carry out targeted actions adjust their courses to learners' requirements. For future research, we wish to embed that tool into an educational recommender system as a module.

## REFERENCES

[1]     A. Bartuskova and O. Krejcar, "Personalization of learning content and its structure, based on limitations of LMSs," in *Studies in Computational Intelligence*, 2016, vol. 642, pp. 455–465.

[2]     D. Verpoorten, C. Glahn, M. Kravcik, S. Ternier, and M. Specht, "Personalisation of learning in virtual learning environments," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009, vol. 5794 LNCS, pp. 52–66.

[3]     M. Ivanović, S. Xinogalos, T. Pitner, and M. Savić, "Technology enhanced learning in programming courses – international perspective," *Educ. Inf. Technol.*, vol. 22, no. 6, pp. 2981–3003, Nov. 2017.

[4]     W. L. Chang, "An agent-based system for collaborative informal learning in a pervasive environment," *Int. J. Mob. Commun.*, vol. 8, no. 2, pp. 187–209, Feb. 2010.

[5]     E. Karataev and V. Zadorozhny, "Adaptive Social Learning Based on Crowdsourcing," *IEEE Trans. Learn. Technol.*, vol. 10, no. 2, pp. 128–139, Apr. 2017.

[6]     N. S. Buot, "Multiple intelligences and reading comprehension of senior high school students: A response evaluation through educational data mining technique," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 6, pp. 2871–2876, Nov. 2019.

[7]     E. Tseytlin, F. Linkov, M. Castine, E. Legowski, and R. S. Jacobson, "Domainbuilder: The knowledge authoring system for slide tutor intelligent tutoring system[version 1; peer review: 1 approved, 1 approved with reservations]," *F1000Research*, vol. 7, 2018.

[8]     S. Schiaffino, P. Garcia, and A. Amandi, "eTeacher: Providing personalized assistance to e-learning students," *Comput. Educ.*, vol. 51, no. 4, pp. 1744–1754, Dec. 2008.

[9]     P. Blessinger, "Creating Meaningful Learning Environments with SLCs," 2015.

[10]    E. J. Castellano and L. MartÃnez, *ORIEB, A CRS FOR ACADEMIC ORIENTATION USING QUALITATIVE ASSESSMENTS*. 2008.

[11]    T. C. Hsia, A. J. Shie, and L. C. Chen, "Course planning of extension education to meet market demand by using data mining techniques - an example of Chinkuo technology university in Taiwan," *Expert Syst. Appl.*, vol. 34, no. 1, pp. 596–602, Jan. 2008.

[12]    C. Vialardi, J. Braver, L. Shaftr, and Á. Ortiaosa, "Recommendation in higher education using data mining techniques," in *EDM'09 - Educational Data Mining 2009: 2nd International Conference on Educational Data Mining*, 2009, pp. 190–199.

[13]    C. Vialardi *et al.*, "A case study: Data mining applied to student enrollment," in *Educational Data Mining 2010 - 3rd International Conference on Educational Data Mining*, 2010, pp. 333–334.

[14]    C. Vialardi *et al.*, "A data mining approach to guide students through the enrollment process based on academic performance," *User Model. User-adapt. Interact.*, vol. 21, no. 1–2, pp. 217–248, Apr. 2011.

[15]    S. Dawson, S. Joksimovic, O. Poquet, and G. Siemens, "Increasing the impact of learning analytics," in *ACM International Conference Proceeding Series*, 2019, pp. 446–455.

[16]    J. Guerra *et al.*, "Adaptation and evaluation of a learning analytics dashboard to improve academic support at three Latin American universities," *Br. J. Educ. Technol.*, Jul. 2020.

[17]    G. Z. Gamboa, "Clustering scholarship programs using educational data mining techniques," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 3, pp. 658–662, May 2019.

[18]    W. Jiang, Z. A. Pardos, and Q. Wei, "Goal-based course recommendation," in *ACM International Conference Proceeding Series*, 2019, pp. 36–45.

[19]    L. C. Freeman, "Centrality in Social Networks Conceptual Clarification," *Soc. Networks*, vol. 1, no. 1968, pp. 215–239, 1978.

[20]    S. P. Borgatti and M. G. Everett, "A Graph-theoretic perspective on centrality," *Soc. Networks*, vol. 28, no. 4, pp. 466–484, Oct. 2006.

[21]    I. Salhi, H. El Fazazi, M. Qbadou, and K. Mansouri, "Towards the Identification of Student Learning Communities using Centrality," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 12.

[22]    I. Salhi and M. Qbadou, "Student learning communities' detection based on betweenness centrality algorithm: Validation and Optimization," in *2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology, IRASET 2020*, 2020.

[23]    G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, Dec. 1966.

[24]    L. C. Freeman, "A Set of Measures of Centrality Based on Betweenness," *Sociometry*, vol. 40, no. 1, p. 35, Mar. 1977.

[25]    U. Brandes, "A faster algorithm for betweenness centrality*," *J. Math. Sociol.*, vol. 25, no. 2, pp. 163–177, Jun. 2001.

[26]    D. A. Keim, M. C. Hao, U. Dayal, H. Janetzko, and P. Bak, "Generalized Scatter Plots," *Inf. Vis.*, vol. 9, no. 4, pp. 301–311, Dec. 2010.

[27]    H. Wickham and L. Stryjewski, "40 years of boxplots," 2011.

[28]    J. W. Tukey, "John W. Tukey - Exploratory Data Analysis-Addison Wesley (1977).pdf." 1977.