11th International Conference on Higher Education Advances (HEAd'25)
Universitat Politècnica de València, València, 2025
DOI: https://doi.org/10.4995/HEAd25.2025.20215

Understanding Student Learning Behaviours in E-Learning: Insights from STEM and Social Science Modules

Le Lai-Hoang, Tai Tan Mai, Martin Crane
School of Computing, Dublin City University, Ireland.

How to cite: Lai-Hoang, L.; Mai, T. T.; Crane, M. (2025). Understanding Student Learning Behaviours in E-Learning: Insights from STEM and Social Science Modules. In: 11th International Conference on Higher Education Advances (HEAd'25). Valencia, 17–20 June 2025. https://doi.org/10.4995/HEAd25.2025.20215

Abstract

The concept of E-learning has gained popularity among universities and students in recent years as E-learning education platforms can record student learning behaviours in its many forms to recognise and analyse student learning styles. However, it is known that the challenges this brings in monitoring student engagement with course material can be considerable and variable between STEM (Science, Technology, Engineering, Math) and Social Sciences courses. This paper applies a graph-based community detection method that integrates the cumulative actions of a student with the Virtual Learning Environment (VLE), through a preprocessing technique, facilitating deeper analysis of student performance using the OULAD dataset[1]. Our findings reveal that this method is trustworthy, and we show that it outperforms traditional classification and clustering methods and achieves superior accuracy in evaluating and predicting academic outcomes—encompassing both formative assignments and terminal assessments. Moreover, this method uncovers variations in learning styles among students in STEM and Social Sciences, indicating commonalities and diversity of the learning approach for the different types of classes, in the same E-learning system.

Keywords: *Learning analytics; E-learning strategies and methodologies; STEM and social sciences; community detection.*

1. Introduction

E-learning platforms are known to create new challenges across both STEM and Social Science modules (Alfaro et al., 2021), requiring educators and students to discover effective learning styles (Downes, 2005). Due to the difficulties in maintaining engagement and communication, managing diverse learning preferences, and integrating technology effectively, having suitable

[1] Open University Learning Analytics Dataset, https://analyse.kmi.open.ac.uk/open_dataset

E-learning strategies is crucial (Ouadoud et al., 2021). Given sufficient E-learning data, Artificial Intelligence (AI) could be expected to be a powerful tool for addressing the learning approach for each kind of module. Additionally, understanding student behaviour in E-learning environments is essential for comprehending how they engage with course materials and how their learning processes are represented (Qiu et al., 2022). In other words, beyond basic course information, E-learning environments offer other meaningful data, such as mouse-clicking behaviour analysis that can be examined in greater depth (Mai et al., 2022). AI can provide insights that contribute to the field of learning analytics (Krugel et al., 2015; Duane, 2024), functioning as a tool to simulate and analyse student E-learning behaviour data. This can enable educators to identify hidden information that informs effective adaptations to learning strategies such as early warning of at-risk students or enhancing course materials quality (Sweeney, 2023). This research enhances the teaching strategies for both STEM and Social Science modules, answering various questions about how both course types benefit from similar methods (Stanigar and Carson, 2020).

In addition, in the domain of applying AI in learning analytics, graph methods are a group of techniques that have shown a significant impact in understanding student behaviours (Porras et al, 2023) and cognitive engagement (Chi et al., 2014). They can represent a group of students in one cohort or class by illustrating their study behaviour data as nodes, and the relationships in learning behaviours as edges, allowing for the recognition of hidden information, such as similarities in learning patterns (Mai et al., 2023). This method has been shown to be highly effective for the visual representation of students in one class (Le et al., 2024), thus potentially offering great benefits to educators (e.g. targeted intervention) in pedagogy. However, considerable effort is needed to preprocess and model data before applying AI methods.

Overall, this paper aims to answer these research questions:

a. How can real-world student learning behaviours be modelled in the context of E-learning?
b. What do these behaviours show in terms of differences in student learning behaviours that can be identified between students in STEM classes and those in Social Sciences classes?

In section 2 of this paper, we apply the method of modelling and analysing the OULAD dataset. After that, we describe the experiment results compared to popular methods and conclude our findings in section 3. Finally, we put our discussion in section 4 before concluding.

2. Methods

2.1. Pre-process method

2.1.1. Feature selection in the OULAD dataset

To reveal commonalities and differences in learning styles between the two types of classes, this research utilised the OULAD dataset (Kuzilek et al., 2017), which features seven e-learning modules in both Social Sciences and STEM that exhibit similar types of online VLE (Virtual Learning Environment) behaviours. The number of interactions for each E-learning element per course per student is counted by date and can be aggregated to describe the learning style at that moment. Consequently, each interaction type is treated as one feature column for the analysis, contributing to the overall learning behaviour. This facilitates a comparison between the two class types—an intriguing topic compared to common questions in this dataset (Jin et al., 2024).

In addition to the VLE interactions, the OULAD dataset records the students' progress assignments and final test results for students of three social sciences and four STEM modules, divided by each class regarding the semester. As both types of modules share similar VLE elements, it allows this research to compare the learning styles of both cases. All the results are converted into two values: 0 for fail and 1 for pass, with 40 as the passing score. Students who deregister during the learning period are excluded, as are assessments with item weights of less than 10%, because they do not greatly add to learning progress. Overall, data from 18,029 e-learners was considered for analytics. Table 1 describes modules information.

Table 1. Information of modules in the OULAD dataset

Module	Number of classes	Number of observed students	Observed formative assignments	Number of types of VLE elements
AAA	2	589	5	9
BBB	3	3609	5	11
CCC	2	2212	4	9
DDD	4	3400	5 or 6	18
EEE	3	1668	4	11
FFF	4	4579	5	18
GGG	3	1972	0	7

2.1.2. Virtual Learning Environment interactions accumulate

The number of interactions with VLE elements is accounted by summation through each action from the time they did the registration, reflecting student engagement in the course and mirroring how teachers summarise data. Afterwards, data is divided into two granularities (one-week or two-week period) for retrospective analysis of student learning behaviours throughout learning time. In addition, for measuring progress, the closest assessment result is applied as the label for a learning interval, along with the terminal test result. For example, if a student has an assignment in the 7th interval, the label for intervals 0 to 7 is that assignment's result. For the next assignment in the 9th interval, its result labels both the 8th and 9th intervals. Moreover, types of VLE interactions are considered as features for the AI method to figure out important factors affecting the test results in modules, which also represent the learning style.

2.2. Graph-based community detection

2.2.1. Graph constructions

Each interval of the learning progress is transformed into a graph, the nodes of which represent student learning behaviour vectors, i.e. the number of VLE features equals the number of dimensions for each node. The graph's edges are calculated based on the correlation between two nodes v_i and v_j via Pearson correlation coefficient (Equation 1) and using the distance metric (Equation 2) to preserve nodal separation after transformation. Therefore, two nodes with short distance imply high similarity of the VLE interactions between two students depicted.

$$C_{ij} = \frac{\sum_{k=1}^{n}(V_{ik}-\overline{V_i})(V_{jk}-\overline{V_j})}{\sqrt{\sum_{k=1}^{n}(V_{ik}-\overline{V_i})^2}\sqrt{\sum_{k=1}^{n}(V_{jk}-\overline{V_j})^2}} \quad (1)$$

$$D_{ij} = \sqrt{0.5 \cdot (1 - C_{ij})} \quad (2)$$

Afterwards, to reduce the complexity of a fully connected graph but preserve the connectivity, a Minimum Spanning Tree (MST) (Graham et al., 1985) is extracted from the original graph to prevent any self-loop in the graph that may cause an unoptimised solution.

2.2.2. Community detection and analysis

A community in a graph can be seen as a group of nodes where members have stronger interactions with each other rather than with external nodes (Fortunato, 2010). This structure can help uncover hidden patterns in a graph, motivating us to adopt it to represent the changes in online learning behaviour patterns among students. Mai et al., 2022 found that students who share the same studying activities tend to have similar learning outcomes. Therefore, it is intuitive to expect that students represented by nodes within the same community might show similar results for both short-term and long-term learning outcomes.

The flow of graph construction and community detection is represented in Figure 1.

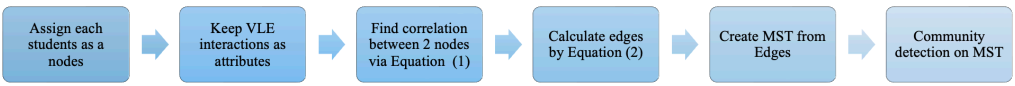

Figure 1. The flow of graph construction and community detection

To find groups of similar students in the graph structure, this paper employs the Louvain method (Blondel et al., 2008), a heuristic method that returns high accuracy with a good time complexity

$O(n \log n)$. Afterwards, a voting process within each community determines the majority label for nodes inside them, leading to a prediction that all students in this community may either pass or fail together. We use the accuracy metrics to verify prediction values and then compare them to the results of five commonly used classification methods: Random Forest (RF), K- Nearest Neighbours (KNN), Support Vector Machine (SVM), XGboost and Logistic Regression (LR) (Jin et al., 2024) to ascertain whether graph-based community detection outperforms other methods to guarantee the analytic results. The key features are identified to reveal the most impactful student learning behaviours, thus highlighting the learning approach for each module before making conclusions on the similarities and differences of STEM and Social Sciences classes in this dataset.

3. Results and discussion

3.1. Community detection and representation

Overall, community detection results outperform those of other classification methods through the learning time, achieving the accuracy about 0.9 across all experiments. For example, Figure 2 shows the accuracy of community detection by the graph method for module DDD, through the 1-week interval, for both formative assignments and terminal result, compared to the results of other methods. Moreover, since the number of communities does not need to be pre-defined, graph methods can significantly outperform traditional clustering techniques on numeric data. Consequently, the result shows a great improvement from other conventional AI-approaches, leading to its reliability for assessing the impact of each feature on the prediction value, thereby highlighting the effect of each VLE interaction on the learning style of each module.

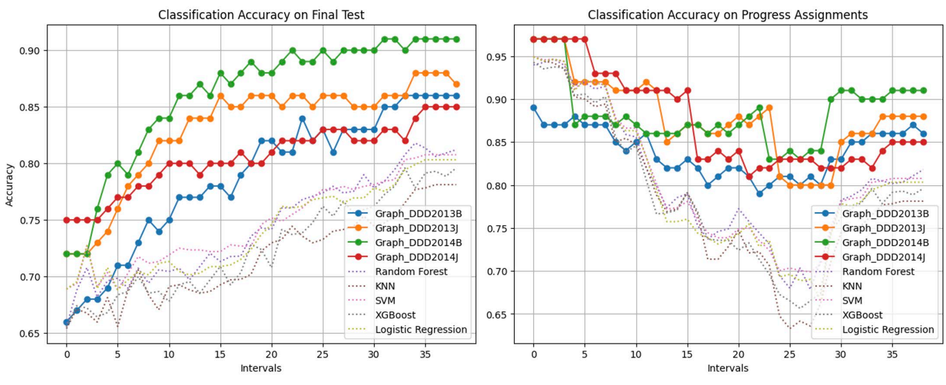

Figure 2. Community detection accuracy for 1-week intervals, versus other methods for 4 DDD modules

Another pedagogical outcome of this method is the representation of student communities for each learning interval, illustrated by the Kamada-Kawai method (Kamada and Kawai, 1988) in Figure 3. The communities detected can facilitate predictions and highlight the communities to which students belong. Educators can utilise these insights to identify groups with similar study behaviours, enabling them to anticipate outcomes more accurately.

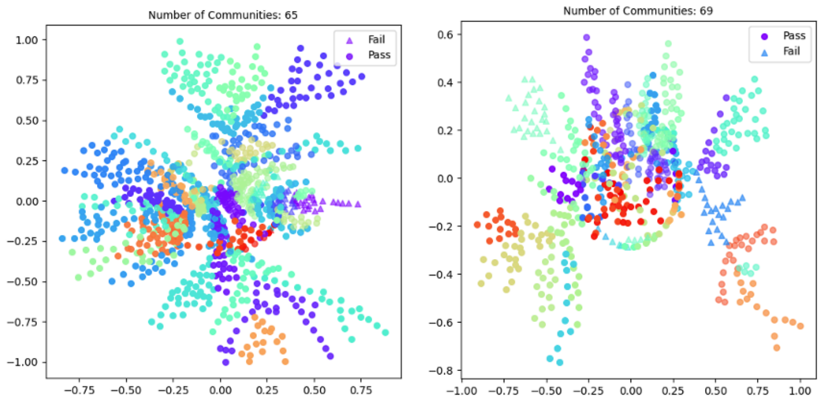

Figure 3. The community visualisation course module FFF, class 2013J in 2 different intervals

3.2. Behavioural interaction analysis for STEM and Social Sciences classes

From the community detection results, we find the most influential VLE elements for each community by using the majority voting and then aggregate them to identify the learning styles for each module. Table 2 presents it, regarding the outcome and comparing between STEM and Social Science types. Overall, the key module elements are "oucontent" and "homepage", indicating that students must engage with the modules' main contents to score well.

Table 2. Most Impactful VLE for Assignments (Bold for Both, Italics for Single Module Type)

Module type	Class	Most influential VLE elements	Most influentialVLE to pass-detected community	Most influential VLE to fail-detected community
Social Scie-nces	AAA	**oucontent**	**oucontent**, subpage, rl, dataplus	oucollaborate, *glossary*
	BBB	**oucontent**, *quiz*, **homepage**	**oucontent,** resources, *quiz*, **homepage**	ouelluminate, oucollaborate, sharedsubpage
	GGG	**oucontent**, *quiz*, forumng	oucontent, forumng, **homepage**, *quiz*	*glossary*, resource, subpage
STEMS	CCC	*subpage*, *forumng*, **homepage**	**homepage**, *subpage* *url*, *forumng*	**oucontent**, resource, oucollaborate
	DDD	*subpage*, resource, wiki	**outcontent,** glossary, wiki *subpage,* externalquiz, *forumng*	page, ouelluminelate, oucollaborate
	EEE	**oucontent**, *url*, **homepage**	**oucontent**, *forumng*, **homepage**, oucollaborate	quiz, page
	FFF	*subpage*, **oucontent**, folder	**homepage**, *subpage*, **oucontent**	htmlactivity, repeatactivity, ouelluminate, oucollaborate

For STEM modules, some VLE elements, e.g., "url," "forumng," and "subpage," significantly contribute to student success. This suggests that seeking information from multiple data sources and participating in forum discussions can lead to better outcomes for students. In contrast, students who focus primarily on resources like "ouelluminelate" or "oucollaborate" tend to struggle with assignments. For Social Science modules, "quiz" is crucial in good result (notably, module AAA lacks the "quiz" VLE), showing the value of completing homework to achieve high scores. Conversely, spending excessive time on elements like "glossary" or "subpage" may indicate that students are not adopting effective strategies for these modules.

4. Discussion and Limitation

Regarding RQ1, our graph-based method effectively models student learning behaviours, grouping students with similar behaviours and academic results into the same communities. In our experiment, it can be seen to outperform traditional methods in predicting academic performance from student learning behaviours. It can provide educators with deeper insights for further analysis. One such analysis is explored in our response to RQ2. We examine STEM modules that require students to conduct extensive research through multiple VLE elements, while Social Sciences modules tend to provide many practices as quizzes for students to improve their study results. Although class content is important for both types of modules, the variances in student learning behaviours show different approaches for students in each type of E-learning module. Overall, with the comparison to other classification methods, this modelling and analytics method shows significant potential application in classroom settings.

This research has limitations related to the dataset. First, student data is aggregated based on community detection for only one interval, which could be improved for better decision-making. Second, extensive data preprocessing and graph formation are needed, making it challenging to apply across classes with varying VLE elements. Third, the absence of other information from classes hinders generalisability, such as interaction with traditional elements like taking note. However, the unique features of OULAD dataset, such as multiple modules with a consistent framework and representation of both STEM and Social Sciences, confirm its reliability for analysis despite being available for several years. In the empirical educational setup, teachers and educators can employ this graph method, combining data from both E-learning environment and static elements to discover the suitable approach for students at every week of the learning period and make interventions for students.

5. Conclusion

In this paper, we proposed a data-driven graph method to model and represent the learning behaviours of students using virtual learning environments in E-learning classes, specifically utilising the OULAD dataset. This method demonstrates outstanding results compared to other approaches and reveals hidden similarities in student learning behaviours across all modules. Based on this modelling, we can detect learning approaches that lead to either good or poor outcomes, highlighting trends in student engagement with specific modules. As a result, it uncovers distinct learning styles for STEM and Social Science classes within this E-learning environment: the former requires considerable effort in researching both the provided course materials and external sources, while the latter emphasises the importance of practising quizzes to achieve desired results. These meaningful insights also suggest that this method has significant potential for application to other learning analytics datasets, offering valuable contributions to pedagogical outcomes such as early detection of "at-risk" students or evaluation of the engagement of students to a specific E-learning object.

Acknowledgements

The author LLH wishes to acknowledge the financial support of the Taighde Éireann – Research Ireland under Grant No. 18/CRT/6223. The authors TTM and MC wish to acknowledge the support in part from the SFI under Grant Agreement No. 13/RC/2106_P2 at the ADAPT SFI Research Centre at DCU.

References

Alfaro, L., Rivera, C., Luna-Urquizo, J., Castañeda, E., Zuniga-Cueva, J., & Rivera-Chavez, M. (2021). New Trends in e-Technologies and e-Learning. In 2021 IEEE World Conference on Engineering Education (EDUNINE) (pp. 1-6). IEEE.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. J. Stat. Mech: Theory Expt, 2008(10), P10008.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. Educational psychologist, 49(4), 219-243.

Downes, S. (2005). E-learning 2.0. ELearn, 2005(10), 1.

Duane, A. (2024). Can artificial intelligence complete my assessment? a student led initiative to stress test the academic integrity of university assessment using generative AI. In 10th Int Conf. on Higher Education Advances (HEAd'24). Editorial Universitat Politècnica de València.

Fortunato, S. (2010). Community detection in graphs. Physics Reports, 486(3-5), 75-174.

Graham, R. L., & Hell, P. (1985). On the history of the minimum spanning tree problem. Annals of the History of Computing, 7(1), 43-57.

Jin, L., Wang, Y., Song, H., & So, H. J. (2024, July). Predictive modelling with the Open University Learning Analytics dataset (OULAD): A systematic literature review. In International Conference on Artificial Intelligence in Education (pp. 477-484). Cham: Springer Nature Switzerland.

Kamada, T., & Kawai, S. (1988). A simple method for computing general position in displaying three-dimensional objects. Computer Vision, Graphics, and Image Processing, 41(1), 43-56.

Krugel, J., Hubwieser, P., Goedicke, M., Striewe, M., Talbot, M., Olbricht, C., ... & Zettler, S. (2020, April). Automated measurement of competencies and generation of feedback in object-oriented programming courses. In 2020 IEEE Global Engineering Education Conference (EDUCON) (pp. 329-338). IEEE.

Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open university learning analytics dataset. Scientific Data, 4(1), 1-8.

Le, L. H., Nguyen, H. D., Crane, M., & Mai, T. T. (2024, June). Multimedia learning analytics feedback in simulation-based training: A brief review. In Proceedings of the 1st ACM Workshop on AI-Powered Q&A Systems for Multimedia (pp. 25-30).

Mai, T. T., Bezbradica, M., & Crane, M. (2022). Learning behaviours data in programming education: Community analysis and outcome prediction with cleaned data. Future Generation Computer Systems, 127, 42-55.

Mai, T. T., Crane, M., & Bezbradica, M. (2023). Students' learning behaviour in programming education analysis: Insights from entropy and community detection. Entropy, 25(8), 1225.

Ouadoud, M., Rida, N., & Chafiq, T. (2021). Overview of E-learning Platforms for Teaching and Learning. Int. J. Recent Contributions Eng. Sci. IT, 9(1), 50-70.

Porras Alfonso, S., Puche Regaliza, J. C., Casado Yusta, S., Sauvée, A., Antón Maraña, P., & Pacheco Bonrostro, J. A. (2023). Relationship between time management and class attendance in university students: clustering techniques for detection of profiles. In 9th International Conference on Higher Education Advances (HEAd'23) (pp. 851-859).

Qiu, F., Zhang, G., Sheng, X., Jiang, L., Zhu, L., Xiang, Q., ... & Chen, P. K. (2022). Predicting students' performance in e-learning using learning process and behaviour data. Scientific Reports, 12(1), 453.

Stanigar, J., & Carson, S. (2020). Adoption of evidenced-based teaching strategies in STEM and non-STEM courses after a common faculty development experience. In 6th International Conference on Higher Education Advances (HEAd'20) (pp. 173-181). Editorial Universitat Politècnica de València.

Sweeney, S. (2023). Academic dishonesty, essay mills, and artificial intelligence: rethinking assessment strategies. In 9th International Conference on Higher Education Advances (HEAd'23) (pp. 19-22).