# Insights from a Socio-Temporal Approach to Student Failure Prediction through Discussion Forum Dynamics

1st Nidia Guadalupe López Flores
*Department of Computer Science*
*Reykjavik University*
Reykjavik, Iceland
nidiaf@ru.is

2nd Víctor Uc-Cetina
*Facultad de Matemáticas*
*Universidad Autónoma de Yucatán*
Mérida, Yucatán
uccetina@correo.uady.mx

3rd Anna Sigridur Islind
*Department of Computer Science*
*Reykjavik University*
Reykjavik, Iceland
islind@ru.is

4th María Óskarsdóttir
*Department of Computer Science*
*Reykjavik University*
Reykjavik, Iceland
mariaoskars@ru.is

***Abstract*—This research paper addresses the significant problem of identifying students-at-risk of failing or dropping out in educational settings. While extensively studied, improving early detection of students likely to fail or drop out remains essential for universities to provide support resources. Previous methods have relied on the students' academic performance to analyse and predict learning strategies, but the complexity of implementing predictive models is heightened by various factors influencing student outcomes. Notably, learning is both a dynamic and socially regulated process, with time and social interactions playing key roles for academic achievement. Nonetheless, despite the importance of these elements, educational research investigating their combined effect is scarce. As grade distribution is affected by several elements, including teaching modalities, grading policies, and course design, identifying students-at-risk and their learning strategies is generally an imbalanced problem, which can lead to biases in predictive algorithms. Our work addresses this issue from a social and temporal perspective, guided by two research questions: (1) To what extent is it possible to inform the early identification of students-at-risk of failing based on interaction data from online discussion forums?, and (2) How does the classification performance compare between traditional oversampling methods and oversampling methods that take the structure of the interactions into account? We based our research on data from an undergraduate course's online forum to build a temporal network of students' communication events across the 12 weeks of the course. Temporal sequences of centrality measures from these interactions served as input for time series classification algorithms. Two oversampling methods are compared: baseline minority oversampling, and a state-of-the-art graph oversampling method that accounts for network structure. Our results show that a temporal network approach, coupled with node oversampling, can enhance student-at-risk identification. However, due to the complexity of the problem and the interactions' sparsity the classification performance is limited when relying solely on this data. We discuss the impact of our findings and contributions, implications, limitations, and future research directions.**

***Index Terms*—students-at-risk, discussion forum, class imbalance, network dynamics, higher education**

## I. Introduction

Identifying students-at-risk (SaR) of failing a course or dropping out of a programme is a significant problem in the fields of Learning Analytics (LA) and Educational Data Mining (EDM) [34]. Their timely detection is critical so that universities and instructors can design and implement effective strategies and interventions to support students. As educational researchers have investigated and identified a large variety of elements associated with and influencing academic failure, multiple approaches and methods have emerged to identify SaR [5]. In addition, educational provision has changed significantly in recent years, introducing new teaching methods, facilitating the use of technology, and providing students with additional resources to support their learning [24]. As a result, in addition to the vast amount of data already available through learning management systems, data from these new technologies, platforms and resources can now be accessed to better inform the identification of SaR.

Categorising students based on their individual strategies, activity patterns, tactics, preferences and personality traits with the objective of understanding the way they learn, and facilitate the design and implementation of interventions to support learners has been extensively addressed by educational researchers (e.g. [19], [28]). However, contrarily to those approaches aiming for predicting dropout from an individual perspective, learning is not entirely an isolated practice or individual achievement [12]. Instead, learning is a socially regulated process [20], where social connections are essential for the development of skills and knowledge, as the students interact, communicate, and receive feedback from other students and instructors [12]. Furthermore, learning is a dynamic process where the effect of time should be explicitly taken into account, as it is not static or instantaneous, but occurs over time [30]. Thus it is essential to consider the effect of time for the timely detection of SaR. As such, although learning should be addressed on both a social and temporal level, approaches combining both aspects have been rarely used by educational researchers [37].

In this paper, we address the imbalanced problem of early identification of SaR of failing the course with a social and temporal perspective. The purpose of our research is to investigate the extent to which concepts of social and temporal analysis are suitable for the identification task, based on data from online discussion forums in undergraduate courses. To this end, two research questions (RQs) guide our work; (1) To what extent is it possible to inform the early identification of SaR of failing based on interaction data from online discussion forums?, and (2) How does the classification performance compare between traditional oversampling methods and oversampling methods that take the structure of the interactions into account?

The participants' interactions on the online discussion forum are used to build the networks. From the temporal perspective, since the progression of students' activity throughout a course has a greater impact on learning than the activity performed during a specific week [38] and the students can be categorised into groups based on their final grade, the methodological approach is multivariate time series classification, which has been effective in detecting early dropout in other contexts as well [49]. We use centrality measures from temporal networks to characterise the relationship between activity in an online discussion forum and academic performance, implement two methods to address the imbalance in the grade distribution, and compare the prediction performance of time series classification models. Regarding the class imbalance, in addition to minority oversampling, we implement the state-of-the-art oversampling method GraphSMOTE [46] to address the grade imbalance at a network level. Thereby, we compute the network measures for the temporal network created and evaluate the performance of both approaches based on multivariate time series classification models. Initial results indicate that although the proposed methodology has the potential to inform the identification of SaR, its effectiveness is negatively influenced by the sparsity in the forum's organic interactions.

## II. Related Work

### A. Academic failure and dropout prediction

Predicting and improving students' academic performance still constitutes one of the main motivations for the LA and EDM adoption in higher education institutions [34], [42]. In this field, analytical tools and procedures are used to understand and improve teaching and learning practices in varied educational contexts [24]. Consequently, the identification of SaR of failing –or dropout– has been addressed with a wide variety of approaches, data sources and features, and data analysis techniques [8]. This task, is usually accomplished by categorising students based on the final grade obtained as a proxy for their academic performance, in order to analyse their learning strategies and predict their academic outcome implementing regression and classification methods [5]. Consequently, researchers have extensively analysed the relationship between failure, academic performance, and activity indices from the students' interactions with learning management systems and other educational platforms [35]. Along with machine learning algorithms, deep learning is increasingly being used in educational research to identify students who need special attention in order to prevent dropout in Massive Open Online Courses (MOOCs) [44], to investigate student behaviour and generate course recommendations [17], and to predict students' academic success [2].

### B. Social and temporal elements of learning

Social Network Analysis has been extensively used to investigate the social component of learning in varied educational contexts [41]. Collecting interaction data can be challenging, time-consuming, and costly. In asynchronous online discussion forums, students interact with other participants and engage in class-related discussions regardless of time and location [22]. In this way, discussion forums data can be analysed from a network perspective [41], with participants as nodes and messages as links [33], to investigate multiple variables, including post activity and interactions [45], topic relevance [45], and student dropout [15].

Learning, like many other real phenomena, is a dynamic practice, as the learning process is not only affected by what happens in the moment. Due to the fact that most educational research has not adequately considered the effect of time on learning activity, the temporal dimension has been underexplored [41]. Such effects relate not only to the moment when the students' activity takes place, but also to the way it happens. Temporal analysis can be used to identify and describe learning events, their variations, metrics, relationships, and transitions [31]. Moreover, undergraduates' self-regulated behaviour and academic performance [40], educational material usage [26], and social connections [33] have been reported to be influenced by time. The study of the temporal element of learning processes and strategies is challenging because it requires measurements over multiple time periods to capture the evolution of its components [39]. A student's approach to learning is also influenced by various factors, including personal characteristics and experience [27]. As affordances between teaching modalities affect students' and teachers' behaviours, it is also essential to consider pedagogical conditions, e.g. course design, technology, and delivery method [14].

### C. Temporal networks and centrality measures

Temporal networks are a particular case of multilayer networks. In these networks, each layer represents the connections (edges) between the same group of entities (nodes) at different time points [29]. Temporal networks extend the concepts of static network analysis to include information on when the interactions between the nodes happen [18]. Centrality measures are understood as numerical indices that characterise multiple features of the relationships between the participants in a network, their position, importance, and influence [13]. In student networks, centrality measures have been useful to investigate the relationship between a student's position in the network and educational outcomes [41]. However, the effect of the centrality measures on academic performance varies depending on several elements, including course setting, participants' background, or the nature of the relationships represented within the networks [36].

Finally, in addition to the challenges in collecting and analysing temporal and social data, identifying SaR of failing is usually an imbalanced problem. The distribution of grades is affected by several elements, including teaching methods and modalities, grading policies, course design, digital ecosystems, and the students' personal characteristics and backgrounds [23], [25]. Consequently, the representation of SaR is not always accurate, which can result in significant biases when predictive algorithms are employed to identify SaR of failing and their learning strategies [23].

## III. Methods

### A. Data and network construction

The data from a discussion forum for an undergraduate course was used to perform the network analysis and prediction of SaR. This course is a required component for first-year students in the computer science programme and second-year students in the engineering programme at the university where the study was performed. The forum involved 323 first- and second-year students, along with 10 instructors, comprised of one teacher and nine teaching assistants.

The teacher accessed the forum data and provided it to us with previous permission from the Teaching Affairs department. When students enrol in this university's academic programmes, they are informed that their data can be used for research and informed consent is required. In this study, only the interactions are used. More specifically, the content of the posts in the discussion forum is neither displayed nor analysed. After matching the users with their grades, all identifiable information was removed. Only the authors had access to the dataset.

The course was taught completely online for 12 weeks. All students enrolled in the course had access to the discussion forum throughout the semester, which was designed to facilitate communication between students and instructors. Active participation was highly encouraged, but not required throughout the course. Therefore, students only posted on the board when they needed or desired, resulting in activity on the forum varying significantly from week to week.

Students received a numerical final grade on a scale of 0 to 10. For the purpose of the analysis, the grades were placed into grade categories based on their numerical grades; A, B, and C corresponded to grades higher than 8.75, 6.75, and 4.75, respectively. Grades under 4.75 were considered failing and were assigned to grade category D. Due to dropping out from the course or using a non-institutional email address to register for the discussion forum platform, some students were assigned to the 'No grade' category. Grade category distribution is A=37 (11%), B=144 (45%), C=83 (26%), D=42 (13%), and No grade=17 (5%).

An example of the participants' interactions in the discussion forum and a demonstration of the way the networks were formed are shown in Fig. 1. The networks are built based on the forum participants' posting activity, reflecting posting and answering behaviour. Directed arrows are drawn from the participants who post replies (source) to those who posted the initial post in each forum thread (target). Additionally, the connecting edges are weighted according to the number of interactions between the nodes. In this study, the data collected do not allow to incorporate reading behaviour in the construction of the networks.

There are several methods to construct networks based on time-dependent communication events. These methods include binary static networks, in which all communications events in a given timeframe are collapsed into one network, and multilayer networks, in which communication events are aggregated based on a specified time window [18]. In our work, a static network was initially constructed with the objective of gaining insights into the overall forum activity and differences in the posting behaviour of students belonging to different grade categories. This static network included all the forum activity during the 12 weeks of the course. This network is presented in Fig. 2 and insights from it are presented in Section IV-A. For the prediction task, a multilayer temporal network was constructed. Considering the asynchronous dynamic and syllabus structure of the course, a time window of one week was used to build the layers. Thus each layer in the temporal network represents communication events between forum participants in each of the twelve weeks of the course. The temporal network is presented in Fig. 3, along with results presented in Section IV-B.

### B. Oversampling methods

There are many factors that influence the distribution of numerical grades in a course. As with other machine learning tasks, unbalanced targets may affect the quality of predictions in detecting SaR of failure. A common method to address class imbalance is to randomly oversample minority classes to prevent biases [10]. Although univariate time series oversampling techniques have been extensively addressed to preserve time dependence, multivariate time series data exhibit additional complexity due to the covariance between time series [47]. Moreover, oversampling time series based on centrality measures should take into account both the links between real and oversampled nodes, since centrality measures depend not only on node attributes, but also on the node's neighbours. To overcome the class imbalance problem, we include two sampling methods, random minority class oversampling and graph synthetic minority oversampling (GraphSMOTE) [46].

The former involves creating multivariate time series of centrality measures from the original temporal network, followed by random oversampling with replacement of minority classes

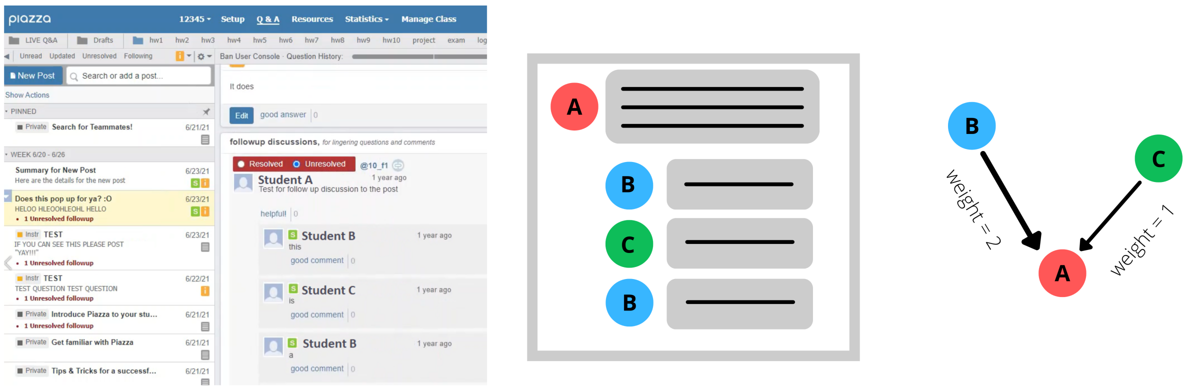


Fig. 1. Left: Discussion forum snapshot example. Right: Network construction example, nodes represent participants in the forum threads, weights are calculated based on the number of interactions between each pair of nodes.

in order to balance the training data and train the time series classifier.

In contrast, the latter generates synthetic nodes for minority classes over an embedding space (an encoder) and models their connections using an edge generator. Afterwards, this information is used to train a graph neural network classifier (decoder). We used Deepwalk [32] and Node2Vec [16] to generate feature representations for the nodes in the network (Table I), and implemented synthetic node generation and link prediction separately for each layer. After parameters optimisation (Table II), the centrality measures were computed using the augmented adjacency matrix created in the latent space. Thus, we compare two sampling strategies, a naive one (random minority oversampling) to a sophisticated one (GraphSMOTE).

Classes A, C, and D were oversampled in both methods. However, the number of synthetic nodes created with GraphSMOTE was fixed to be less than the number of nodes of each category in the training set (80%). This was decided to prevent the creation of nodes with redundant information [46].

### C. Time series classification

In this study, sequences of centrality measures for the student nodes were used as input for the time series classification algorithms. The centralities' selection was based on previous educational research. For instance, [36] explored the role of centrality measures as indicators of academic success based on data from the online collaboration tasks, showing degree centralities and eigenvector centrality were consistent indicators of academic performance. Furthermore, [41] reported that although the study of network centralities in education is extensive, most research is limited to traditional connectivity measures, e.g. betweenness, closeness and degree; highlighting that the inclusion of novel centralities, e.g. Katz centrality, would inform underexplored elements of learning processes. Table III lists the centrality measures selected along with their descriptions.

All the centrality measures listed in Table III were used together as input for the multivariate time series classification models. The models were implemented and evaluated by concatenating the centralities across various layers, resulting in a 3-dimensional input array with dimensions (number_of_students) x (number_of_centralities) x (number_of_weeks). Five incremental time horizons were considered: four, six, eight, ten, and twelve weeks. For each timeframe, the three node feature representations in Table I were used.

This paper primarily focuses on identifying categories A to D. However, 'No grade' students and instructors are also included in the networks' nodes oversampling and centralities computation as they are essential for the information dynamics [25]. Both oversampling techniques were assessed by randomly selecting 20% of each class from the original dataset and setting it aside to ensure that synthetic nodes were not included in the test set.

Regarding the multivariate time series classification models, Rocket [9] and K-neighbours [7] time series classifiers were pre-trained and 10-fold cross-validation was used to find the optimal parameters (Table IV). We evaluated how well the model could differentiate between classes using the Area Under Receiver Operator Curve (AUC) score [4]. For multiclass classification, the weighted average AUC score was calculated for each class against the rest [11].

## IV. Results

### A. Forum network description

Figure 2 displays the static network constructed considering all the interaction events that occurred among the students throughout the 12 weeks of the course. As described in Section III-A, students are represented by nodes in the network, which are coloured in the figure according to the grade category the students were assigned to, based on the final grade obtained in the course. Teacher and teaching assistants are also included in the network with nodes coloured yellow and brown, respectively. Further, thicker edges indicate a higher number of interactions between the nodes they connect. In contrast, thinner edges indicate weaker, less frequent interactions throughout the course. Additionally, in Figure 2, the nodes'

TABLE I
PARAMETERS LIST FOR FEATURE REPRESENTATION OF NODES IN THE NETWORK

| Node feature representation | Parameters |
|---|---|
| Deepwalk | walk_number=10, walk_length=80, dimensions=356, workers=4, window_size=10 |
| Node2Vec Explore | walk_number=10, walk_length=80, p=2, q=0.5, dimensions=356, window_size=10 |
| Node2Vec Stay Locally | walk_number=10, walk_length=80, p=0.5, q=2, dimensions=356, window_size=10 |

TABLE II
PARAMETER GRID SEARCH FOR THE NODE OVERSAMPLING WITH GRAPHSMOTE

| Parameter | Grid search space |
|---|---|
| model | {'sage', 'GAT'} |
| nhid (hidden layers) | {64,128,256} |
| lr | {0.001, 0.01} |
| dropout | {0.2, 0.5, 0.8} |

sizes are proportional to the number of posts each participant made in the discussion forum. Larger nodes correspond to participants who were more active and posted more frequently, while smaller nodes indicate participants who were less active throughout the course. These features provided insights into the forum activity. Firstly, it is noticeable the key role played by the teacher and teaching assistants in supporting students by posting answers to the students' questions in the forum. In addition, the nodes' sizes also show differences between grade categories. Red and blue nodes are significantly larger than green and purple nodes (Mann Whitney U-Statistic = 14123.5, P-value= 6.2x$10^{-08}$), indicating that students in A and B grade categories were more active than students in C and D grade categories. Further, in addition to posting behaviour, there are differences in the extent to which students in each grade category are connected to other nodes within the network. The average degree, as described in Table III, for each of the four grade categories are, A=7.03, B=4.15, C=2.53, and D=0.95. These values indicate students in grade category A interact, on average, with more students in comparison with students in other categories. These insights into the differences in the nodes' posting activity and their connectivity within the forum suggest that analysing these interaction patterns can be a helpful approach for identifying SaR.

Figure 3 displays the temporal network build based on the 12 weeks of interaction as described in Section III-A. The network comprises 12 layers, each corresponding to the activity that took place during the respective week. Considering both the variations in the activity between grade categories and the differences in the posting activity within weeks, the centrality measures calculated from the temporal network's layers are helpful indicators of changes in activity levels and connectivity throughout the course.

### *B. Oversampling and identification of SaR*

The effects of the graph synthetic minority oversampling over the grade distribution were assessed by calculating the imbalance ratio $(I_r)$ [46] between grade categories. This ratio was computed to compare the ratio of failed students before and after applying GraphSMOTE in comparison with

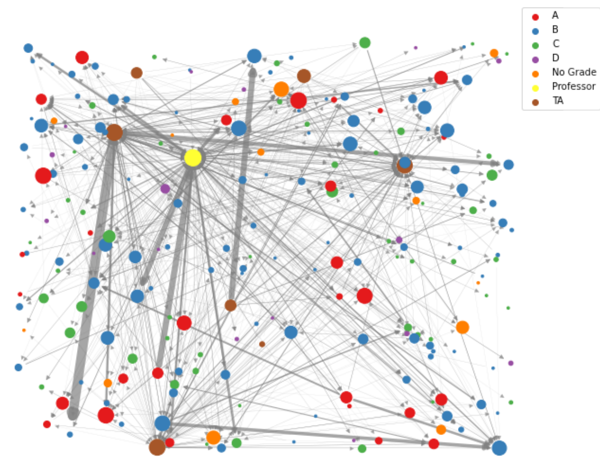


Fig. 2. Forum network built considering all the interaction events that occurred in the discussion forum threads during the 12 weeks of the course. Nodes are coloured based on grade categories, A (red), B (blue), C (green), D (purple), No grade (orange). The instructors, namely teacher and teaching assistants are represented by yellow and brown nodes, respectively.

the other grade categories. The $I_r$ was adapted for courses where the D grade category does not necessarily correspond to the smallest class, defining it as the number of failed students (category D) over the number of students in the most frequent category (category B). Results of the GraphSMOTE oversampling effects in the class distribution are displayed in Equations (1) and (2).

$$I_{r\ original\ distribution} = \frac{(Class\ D)}{(Class\ B)} = 0.29 \quad (1)$$

$$I_{r\ GraphSmote} = \frac{(Class\ D)}{(Class\ B)} = 0.44 \quad (2)$$

As described in Section III-C, for each timeframe and oversampling method, the temporal sequences were created by concatenating the centrality measures across weeks, evaluating the models on Table (IV). Table V presents the results of the classification task for both oversampling methods across the five timeframes. For random oversampling of minority classes, it highlights the models that achieved the highest training AUC scores during the training phase. Meanwhile, for synthetic oversampling using GraphSMOTE, it displays the node feature representations and models with the highest AUC scores for each timeframe, along with their performance outcomes on both the training and test sets. These results clearly demonstrate that the models based on centrality sequences using GraphSMOTE across the five timeframes consistently outperformed those that utilized random oversampling of minority

TABLE III
CENTRALITY MEASURES USED AS INPUT FOR TIME SERIES CLASSIFICATION MODELS. SELECTION BASED ON [36], [41].

| Measure | Description |
|---|---|
| Degree | For a node v, is the fraction of nodes it is connected to. |
| In-degree | For a node v, is the fraction of nodes its incoming edges are connected to. |
| Out-degree | For a node v, is the fraction of nodes its outgoing edges are connected to. |
| Betweenness | For node v, is the sum of the fraction of all-pairs shortest paths that pass through v. |
| Closeness | For a node v, is the reciprocal of the average shortest path distance to all other reachable nodes in the network. Higher values indicate better connectedness. |
| Eigenvector | For a node v, it quantifies its influence in the network by assigning a score obtained through iterative aggregation of the centrality of its connected neighbours. Connections to highly central nodes contribute more to this score. |
| Katz | Computes the relative influence of a node v by measuring the number of immediate neighbours and all other nodes that connect to v through them. |
| Page rank | Computes a ranking of the nodes in the graph based on the structure of the incoming links. |

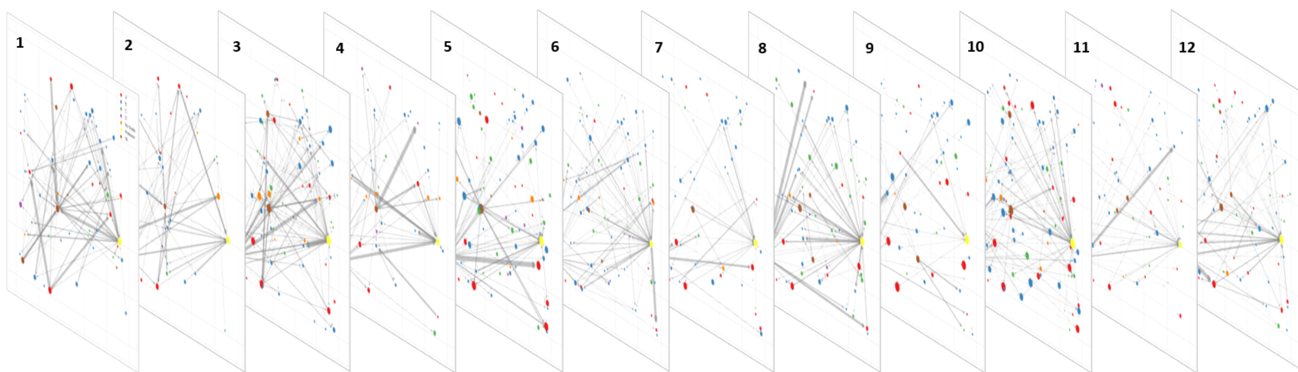


Fig. 3. Temporal network built based on the 12 weeks of interaction events. Each layer represents the interaction events that occurred in each of the 12 weeks of the course. Nodes are coloured based on grade categories, A (red), B (blue), C (green), D (purple), No grade (orange). Instructors, are represented by yellow and brown nodes.

TABLE IV
PARAMETER GRID SEARCH FOR THE MULTIVARIATE TIME SERIES CLASSIFICATION

| Model | Grid search space |
|---|---|
| Rocket | num_kernels: {1000, 2000, 3000} |
| K-neighbours TS | distance: {‘dtw’, ‘euclidean’, ‘squared’} |

classes, achieving higher AUC scores. Nonetheless, despite achieving the highest scores with GraphSMOTE across the five timeframes, the test set performance reveals significant gaps between training and testing results. This suggests that while the models benefited from oversampling via GraphSMOTE combined with Node2Vec (Explore) and DeepWalk node feature representations, their overall performance is still not optimal for the identification of SaR and it is likely impacted by underlying biases within the dataset.

These results suggest that the synthetic oversampling method, along with the implemented classification models, could be effective in identifying SaR and shedding light on the centrality measures useful for this purpose. However, the practical application of these methods may be constrained by the complexity, sparsity, and potential biases in the data.

## V. DISCUSSION

The timely identification of SaR is a significant and challenging task. To achieve this goal, the identification of valuable data sources and selection of relevant features is critical. By using online discussion forums, where students voluntarily post questions and answer other students' questions, we assess whether pure interaction data from online discussion forums is useful in identifying SaR. Our approach does not depend on students' background information, course assessment elements, or the content of the comments posted on the discussion forum. We implement multivariate time series classification models using interaction data and network science concepts to identify students in various grade categories across multiple time points. Moreover, our research approach also takes into consideration the imbalanced nature of grade categories.

Our research results can be analysed and discussed from multiple perspectives. Firstly, in regard to the nature of the problem we aim to address. As presented in Section II-A, the identification of SaR is a research topic that has been extensively addressed by several communities, including educational research, LA, and EDM. However, although the accuracy in the prediction of SaR has improved significantly with the development of intelligent systems, such as machine and deep learning models, the implementation in real-life learning settings is still limited. There is a growing need of bridging the gap between the identification of SaR, and effectively

TABLE V

COMPARISON OF CLASSIFICATION MODEL PERFORMANCE USING RANDOM OVERSAMPLING (LEFT) VERSUS GRAPH SYNTHETIC OVERSAMPLING WITH GRAPHSMOTE OVER DIFFERENT TIMEFRAMES (RIGHT). GRAPHSMOTE-BASED MODELS WERE EVALUATED ON THE TEST SET; AUC (TEST), AND $AUC_D$ (TEST) FOR THE FAILED STUDENTS CLASS ARE DISPLAYED.

| | **Random oversampling** | |
|---|---|---|
| **Weeks** | **Model** | **AUC(training)** |
| 1 to 4 | Rocket | 0.5342 |
| 1 to 6 | K-neighbours | 0.5310 |
| 1 to 8 | Rocket | 0.5591 |
| 1 to 10 | K-neighbours | 0.5506 |
| 1 to 12 | K-neighbours | 0.5562 |

| **Graph synthetic oversampling with GraphSMOTE** | | | | |
|---|---|---|---|---|
| **N.f. representation** | **Model** | **AUC(training)** | **AUC(test)** | **AUC$_D$(test)** |
| Node2Vec Explore | Rocket | 0.5695 | 0.5360 | 0.4906 |
| Node2Vec Explore | K-neighbours | 0.5716 | 0.4739 | 0.4203 |
| Node2Vec Explore | K-neighbours | 0.5873 | 0.5012 | 0.5042 |
| DeepWalk | K-neighbours | 0.5873 | 0.4194 | 0.3658 |
| DeepWalk | K-neighbours | 0.6100 | 0.4106 | 0.3753 |

preventing their their failure. Addressing these challenges is crucial; the identification of SaR is not only key to prevention but also vital for implementing and evaluating models and methodologies in real-world settings. This implementation could lead to the identification of failure-related actions and behaviours, thereby facilitating the development of resources to support effective learning.

In regards to the research approach selected, while both the social learning [3] and self-regulation theories [43] attest to the importance of social and temporal aspects, research that integrates both elements is scarce [37]. Prior research addressing these elements has been conducted in diverse educational settings, including MOOCs, online, and distance learning. However, although most higher education programmes have online or hybrid elements, prior research cannot be extended to existing higher education environments, as changes in the teaching modality lead to changes in the way the students use the learning platforms and resources [24], [25]. Our work contributes to the field by adopting an approach that recognises the joint influence of social and temporal elements on learning processes. Moreover, we focus our analysis on interaction data from an online discussion forum where the students' activity only includes organic contributions, which is common in higher education courses. Thus, the weekly activity was truly sparse since participation in discussion threads was neither mandatory nor part of the assessment structure.

The initial results obtained have several implications. Regarding the first RQ presented in Section I: 'To what extent is it possible to inform the early identification of SaR of failing based on interaction data from online discussion forums?' the results of classification models based on random oversampling show that even though prior research in other settings has found this type of data useful to identify SaR [15], [44], this approach does not have the same direct benefit when the data come from courses where the structure is different and the students' participation on the forum is organic. Regarding the second RQ: 'How does the classification performance compare between traditional oversampling methods and oversampling methods that take the structure of the interactions into account?', our analysis showed that regardless of timeframe, the average AUC was higher when oversampling was performed with GraphSMOTE for training. However, as shown in Section IV, for four out of five timeframes, the best classification models performed below 0.5 (random baseline) over the test set ($AUC_D$, Table V). There are several reasons for this effect, including the fact that students belonging to each category differ in their activity levels [26], [40]. Activity data are therefore not only imbalanced, but also limited in terms of identifying activity patterns, in line with research on minorities' visibility on social networks, showing that increasing their visibility is not only influenced by group size, but also by group behaviour [6]. In practice, encouraging the students to be more active in the forum would benefit both students as well as the models' implementation and performance.

Our work has several limitations and offers many possibilities for further research. The first limitation relates to the data itself. As we have extensively stressed, the dataset is truly sparse throughout the course duration, implying both the class imbalance and low activity levels have an impact on the performance of the classification models. However, we consider this limitation as one of our main contributions, as it reflects the data structure of several computer science and engineering undergraduate programmes. Therefore, research on how to make the most of this data is highly important to facilitate the higher education institutions adoption of data-driven analyses developed to improve teaching and learning practices. Another limitation relates to the course selected and its teaching modality. In our work, although fully onsite and hybrid courses were available, we decided to rely on a course that was taught completely online, as it was the course with more activity in the online discussion forum. Nevertheless, previous research has demonstrated that the teaching modality selected affects the interaction dynamics in online discussion forums [25], limiting the conclusions of this research, which may not apply to courses that adopt a different teaching modality. It is therefore recommended to extend our work to evaluate results in courses belonging to different settings, including teaching modality and assessment structure. Finally, although our approach allowed us to improve the classification performance against the traditional approach performed with random oversampling the minority classes, the low classification performance obtained is seen as the final limitation of this work. The performance score indicates that even though the performance was improved by our method, interaction data alone might not provide enough information to fully identify SaR. Nonetheless, as network data on its own tend to be limited [48], we consider our approach could be used to boost the classification performance of more complex models that combine network data with data from other educational platforms.

Although analysing this type of data may have potential benefits, it also raises ethical and privacy concerns [1]. It is true that LA research can help develop tailored learning experiences for students and provide adequate resources to support effective learning [1], but it is also true that students who are identified as ‘at-risk’ might feel discouraged, therefore having the opposite effect [21]. These ethical and privacy concerns should also be taken into consideration when implementing EDM and LA research results, requiring the development of institutional policies ensuring appropriate data collection, processing, and models’ implementation [1]. To this end, it is crucial to explore data coming from educational platforms to identify data sources that can inform the early identification of SaR.

Future work of this research will focus on addressing the limitations described. The approach will be tested in other courses and educational settings, such as different subjects, teaching modalities, and courses with graded forum participation. Additionally, data from other sources such as learning materials usage and system logs will be incorporated to evaluate the extent to which forum interactions enhance classification models. Lastly, future work will explore the observed performance differences between training and testing phases, as well as investigate alternative classification models for identifying SaR.

In summary, from a social and temporal perspective, our work explores the extent to which pure interaction data from online discussion forums can be used to identify SaR. On the one hand, our findings provide insights into the forum activity, highlighting significant differences between students belonging to different grade categories. On the other hand, the comparison between the two oversampling methods implemented to evaluate the predictive power of forum data indicates that classification performance can be improved by accounting for the networks’ structure. Nonetheless, our findings also call attention to the characteristics and limitations of the data, further extending the implications to the need to improve and encourage student participation in discussion forums.

## VI. Conclusion

The identification of SaR has been extensively addressed in the past by LA and EDM researchers. Nevertheless, more research is needed to properly account for the effect of social and temporal elements on learning processes in different educational environments. Our study takes a further step on the identification of SaR from a joint social and temporal perspective based on network measures and multivariate time series classification models. We rely on the limited and sparse data from online discussion forums in undergraduate courses taught to computer science and engineering undergraduate students to evaluate the extent to which our approach enhances classification models aiming for identifying SaR at early stages.

The main contribution is two-fold: (i) lessons learnt from exploration of early identification of SaR of failing and (ii) insights into ways in which classification performance compare between traditional oversampling methods and oversampling methods that take the structure of the interactions into account. We show the approach adopted has the potential to improve the classification performance at different timepoints, but such improvement, and in consequence its implementation, are limited by the sparsity in the students’ interactions in the forum. Future work will extend to incorporating data from other educational platforms, testing our methodology across various educational settings, and exploring a broader range of classification algorithms.


## Acknowledgment

The first author acknowledges the support of the Icelandic Research Fund (IRF) [grant number 239408-051].


## References


[1] S. J. Aguilar. Learning analytics: at the nexus of big data, digital innovation, and social justice in education. *TechTrends*, 62:37–45, 2018.

[2] M. Akour, H. A. Sghaier, and O. Al Qasem. The effectiveness of using deep learning algorithms in predicting students achievements. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(1):388, Jul 2020.

[3] A. Bandura. Social learning through imitation. 1962.

[4] A. P. Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, Jul 1997.

[5] C. Brooks and C. Thompson. *Predictive Modelling in Teaching and Learning*, page 29–37. SoLAR, Vancouver, Canada, 2 edition, 2022.

[6] Z. Budrikis. Intervening in networks to make minorities more visible. *Nature Reviews Physics*, 5(8):445–445, Aug. 2023.

[7] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, Jan. 1967.

[8] G. Deeva, J. D. Smedt, C. Saint-Pierre, R. Weber, and J. D. Weerdt. Predicting student performance using sequence classification with time-based windows. *Expert Systems with Applications*, 209:118182, 2022.

[9] A. Dempster, F. Petitjean, and G. I. Webb. Rocket: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*, 34(5):1454–1495, 2020.

[10] M. R. K. Dhurjad and M. S. S. Banait. Imbalanced time series data classification using over-sampling technique. pages 75–80, 2015.

[11] T. Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874, Jun 2006.

[12] R. Ferguson and S. B. Shum. Social learning analytics. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK ’12*, pages 23–33. ACM Press, 2012.

[13] N. E. Friedkin. Theoretical foundations for centrality measures. *American Journal of Sociology*, 96(6):1478–1504, May 1991.

[14] D. Gašević, S. Dawson, T. Rogers, and D. Gasevic. Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, 28:68–84, 2016.

[15] N. Gitinabard, F. Khoshnevisan, C. F. Lynch, and E. Y. Wang. Your actions or your associates? predicting certification and dropout in moocs with behavioral and social features. In *International Educational Data Mining Society*, 2018.

[16] A. Grover and J. Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 855–864, San Francisco California USA, Aug 2016. ACM.

[17] A. Hernández-Blanco, B. Herrera-Flores, D. Tomás, and B. Navarro-Colorado. A systematic review of deep learning approaches to educational data mining. *Complexity*, 2019:1–22, May 2019.

[18] P. Holme and J. Saramäki. *Temporal network theory*, volume 2. Springer, 2019.

[19] J. Jovanović, S. Dawson, S. Joksimović, and G. Siemens. Supporting actionable intelligence: reframing the analysis of observed study strategies. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, page 161–170, 2020.
[20] S. Järvelä, P. A. Kirschner, E. Panadero, J. Malmberg, C. Phielix, J. Jaspers, M. Koivuniemi, and H. Jär-venoja. Enhancing socially shared regulation in collaborative learning groups: Designing for cscl regulation tools. *Educational Technology Research and Development*, 63(1):125–142, 2015.
[21] C. Lawson, C. Beer, D. Rossi, T. Moore, and J. Fleming. Identification of 'at risk' students using learning analytics: the ethical dilemmas of intervention strategies in a higher education institution. *Educational Technology Research and Development*, 64(5):957–968, Oct 2016.
[22] D. Lee, R. Rothstein, A. Dunford, E. Berger, J. F. Rhoads, and J. DeBoer. "connecting online": The structure and content of students' asynchronous online networks in a blended engineering class. *Computers & Education*, 163:104082, 2021.
[23] O. H. Lu, A. Y. Huang, and S. J. Yang. Impact of teachers' grading policy on the identification of at-risk students in learning analytics. *Computers & Education*, 163:104109, Apr 2021.
[24] N. López Flores, A. S. Islind, and M. Óskarsdóttir. *A learning analytics-driven intervention to support students' learning activity and experiences*, page 81–102. Routledge, London, 1 edition, Oct. 2023.
[25] N. López Flores, M. Óskarsdóttir, and A. S. Islind. Analysis of discussion forum interactions for different teaching modalities based on temporal social networks. In *Proceedings of the NetSciLA22 workshop, March 22, 2022*, pages 23–32, 2022.
[26] N. G. López Flores, A. S. Islind, and M. Óskarsdóttir. Making the most of slides and lecture captures for better performance: A learning analytics case study in higher education. In *The 56th Hawaii International Conference on System Sciences (HICSS)*, pages 1291–1300, 2023.
[27] P. Marriott. A longitudinal study of undergraduate accounting students' learning style preferences at two uk universities. *Accounting Educa-tion*, 11(1):43–62, 2002.
[28] W. Matcha, D. Gašević, N. A. Uzir, J. Jovanović, and A. Pardo. Analytics of learning strategies: Associations with academic performance and feedback. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, page 461–470, 2019.
[29] F. Menczer, S. Fortunato, and C. A. Davis. *A first course in network science*. Cambridge University Press, 1 edition, 2020.
[30] N. Mercer. The seeds of time: Why classroom dialogue needs a temporal analysis. *The journal of the learning sciences*, 17(1):33–59, 2008.
[31] I. Molenaar and A. F. Wise. *Temporal As-pects of Learning Analytics - Grounding Analyses in Concepts of Time*, page 66–76. SoLAR, Vancouver, Canada, 2 edition, 2022.
[32] B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk: online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, page 701–710, New York New York USA, Aug 2014. ACM.
[33] O. Poquet. Why birds of a feather flock together: Factors triaging students in online forums. In *LAK21: 11th International Learning Analytics and Knowledge Conference*, page 469–474, 2021.
[34] C. Romero and S. Ventura. Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3):e1355, May 2020.
[35] J. Russell, A. Smith, and R. Larsen. Elements of success: Supporting at-risk student resilience through learning analytics. *Computers & Education*, 152:103890, Jul 2020.
[36] M. Saqr and S. López-Pernas. The curious case of centrality measures: A large-scale empirical investigation. *Journal of Learning Analytics*, 9(1):13–31, 2022.
[37] M. Saqr and S. López-Pernas. *The Why, the What and the How to Model a Dynamic Relational Learning Process with Temporal Networks*. Nov 2022.
[38] M. Saqr, S. López-Pernas, S. Helske, and S. Hrastinski. The longitudinal association between engagement and achievement varies by time, students' profiles, and achievement state: A full program study. *Computers & Education*, 199:104787, July 2023.
[39] M. Saqr, S. López-Pernas, J. Jovanović, and D. Gašević. Intense, turbulent, or wallowing in the mire: A longitudinal study of cross-course online tactics, strategies, and trajectories. *The Internet and Higher Education*, 57:100902, Apr 2023.
[40] M. Saqr, J. Nouri, and U. Fors. Temporality matters. a learning analytics study of the patterns of interactions and its relation to performance. In *EDULEARN18 Proceedings*, 10th International Conference on Education and New Learning Technologies, page 5386–5393, Palma, Spain, Jul 2018. IATED.
[41] M. Saqr, O. Poquet, and S. López-Pernas. Networks in education: A travelogue through five decades. *IEEE Access*, 10:32361–32380, 2022.
[42] Y.-S. Tsai, D. Rates, P. M. Moreno-Marcos, P. J. Muñoz-Merino, I. Jivet, M. Scheffel, H. Drachsler, C. Delgado Kloos, and D. Gašević. Learning analytics in european higher education—trends and barriers, 2020.
[43] P. H. Winne. *Learning Analytics for Self-Regulated Learning*, page 78–85. SOLAR, 2 edition, 2022.
[44] W. Xing and D. Du. Dropout prediction in moocs: Using deep learning for personalized intervention. *Journal of Educational Computing Research*, 57(3):547–570, Jun 2019.
[45] B. Yang, H. Tang, L. Hao, and J. Rose. Untangling chaos in discussion forums: A temporal analysis of topic-relevant forum posts in moocs. *Computers & Education*, 178:104402, Jul 2021.
[46] T. Zhao, X. Zhang, and S. Wang. Graphsmote: Imbalanced node classification on graphs with graph neural networks. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, page 833–841, Virtual Event Israel, Mar 2021. ACM.
[47] T. Zhu, C. Luo, Z. Zhang, J. Li, S. Ren, and Y. Zeng. Minority oversampling for imbalanced time series classification. *Knowledge-Based Systems*, 247:108764, Jul 2022.
[48] M. Óskarsdóttir, C. Bravo, C. Sarraute, J. Vanthienen, and B. Baesens. The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics. *Applied Soft Computing*, 74:26–39, Jan. 2019.
[49] M. Óskarsdóttir, T. Van Calster, B. Baesens, W. Lemahieu, and J. Vanthienen. Time series for early churn detection: Using similarity based classification for dynamic networks. *Expert Systems with Applications*, 106:55–65, Sep 2018.