How learners' interactions sustain engagement: a MOOC case study

Ayse Saliha Sunar, Su White, Nor Aniza Abdullah, and Hugh C. Davis

Abstract

In 2015, 35 million learners participated online in 4,200 MOOCs organised by over 500 universities. Learning designers orchestrate MOOC content to engage learners at scale and retain interest by carefully mixing videos, lectures, readings, quizzes, and discussions. Universally, far fewer people actually participate in MOOCs than originally sign up with a steady attrition as courses progress. Studies have correlated social engagement to completion rates. The FutureLearn MOOC platform specifically provides opportunities to share opinions and to reflect by *posting* comments, *replying*, or *following* discussion threads. This paper investigates learners' social behaviours in MOOCs and the impact of engagement on course completion. A preliminary study suggested that dropout rates will be lower when learners engage in repeated and frequent social interactions. We subsequently reviewed the literature of prediction models and applied social network analysis techniques to characterise participants' online interactions examining implications for participant achievements. We analysed discussions in an eight week FutureLearn MOOC, with 9855 enrolled learners. Findings indicate that if learners starts following some , the probability of their finishing the course is increased; if learners also interact with those they follow, they are highly likely to complete, both important factors to add to the prediction of completion model.

Index Terms

Social network analysis, MOOC, peer interactions, prediction model, learning at scale

1 INTRODUCTION

EDUCATIONAL technologists see the web as a gigantic information system that reveals learning resources to a growing world population for the purpose of creating richer learning experiences. Widespread use of the web in teaching and learning has resulted in the emergence of new pedagogies and learning paradigms in recent years. In 2005, for example, Siemens proposed a new "connectivist" learning theory for the digital age [1].

Via connectivism, people learn by making meaningful connections amongst knowledge, information resources and ideas during the learning process. Web and social technologies facilitate the acquisition of useful knowledge and establish cognitive connections. Hailed as the first MOOC, Connectivism and Connective Knowledge'08 (CCK'08) was based on Siemens' learning theory. It was launched in Autumn 2008 with 2000 people showing initial interest. In order to stimulate communication during the course, video lectures and tasks were regularly released, and group discussions are encouraged amongst participants via popular social media platforms, wikis and blogs.

The title MOOC (Massive Open Online Course) is coined later [2]. Subsequently, numerous universities and private institutions have launched their own MOOCs and in turn, thousands of learners now participate in the courses[1].

Not all MOOCs are crafted around connectivism. Rodriguez found that cognitive-behaviourist and social constructivist approaches have also been adopted by MOOCs [3].

- *Ayse Saliha Sunar, Su White and Hugh C Davis are with the Department of Electronics and Computer Science, University of Southampton, United Kingdom.*
  E-mail: ass1a12@soton.ac.uk
- *Nor Aniza Abdullah is with the University of Malaya.*
  E-mail: noraniza@um.edu.my

Manuscript received February 1, 2016; revised August XX, 2016.

1. https://www.class-central.com/report/moocs-2015-stats

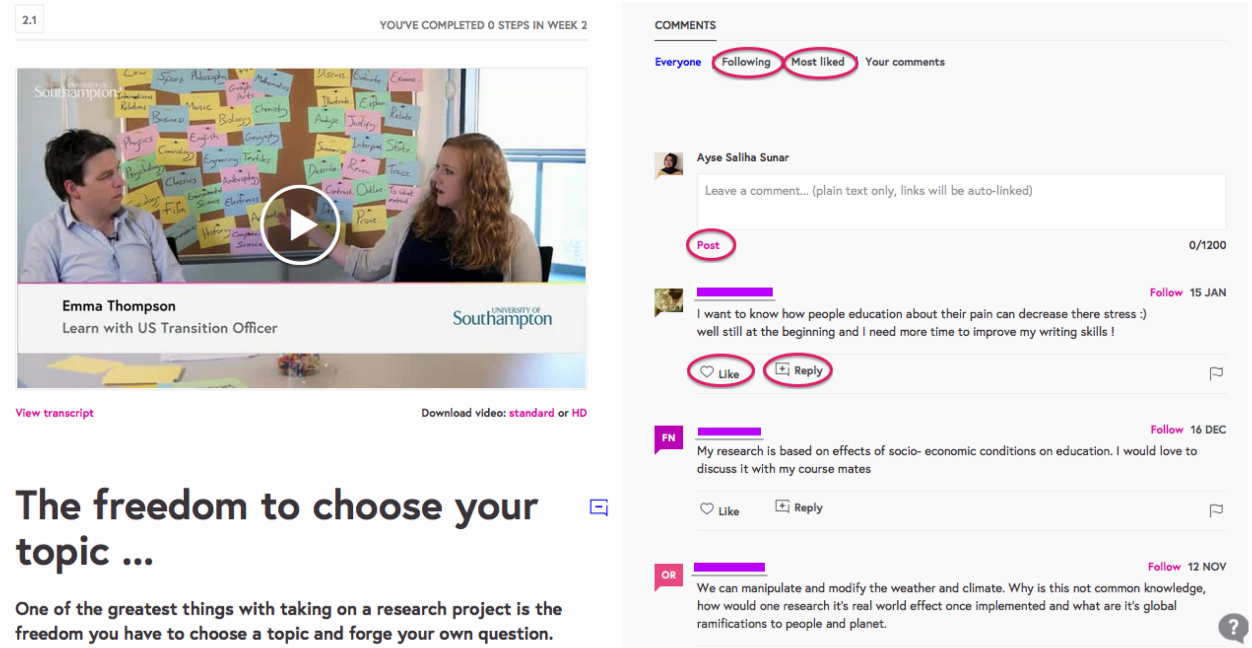


Fig. 1. The FutureLearn platform, highlighting social affordances within discussion thread.

Two celebrity professors from the Stanford University, Sebastian Thrun and Peter Norvig, launched their CS221 MOOC: Introduction to Artificial Intelligence (AI-Stanford) based on their classroom teaching, it was taught online alongside face-to-face delivery during 2011. Shortly after, Thrun and Norvig launched the MOOC platform, Udacity, that now offers many other free online courses mainly on technical subjects[2].

Courses developed using Udacity, and other similarly featured MOOC platforms such as EdX and Coursera are based on cognitive-behaviourist pedagogy with some small components from social constructivism. These are more centralised than the distributed connectivist model pioneered by Siemens [4]. In these later frameworks, the role of instructors is similar to face-to-face teaching and the lectures are structured so that each week has a defined set of learning objectives [3]. Consequently, interactivity amongst learners and instructors is limited. In cognitive behaviourist MOOCs, social tools integrated for communication via forums and linked Facebook groups when they are mostly used for asking questions about course content and assessments rather than the co-creation and co-evolution of learning content of connectivist MOOCs [3].

Researchers have discussed the most suitable pedagogy for MOOCs to enable learners to have the best possible online learning experience (e.g. [5], [6], [7], [8]). FutureLearn, the UK-based MOOC platform owned by The Open University, is seeking to develop a pedagogy that works at massive scale [8]. The design team led by Professor Mike Sharples has implemented a social-constructivist learning theory based on Laurillard's conversational framework [9]. The FutureLearn platform is designed to promote successful conversations providing participants with links between the visible repository of learning resources and a set of integrated tools which enable commenting, responding and reflection (see Section 2) [8].

1.1 Participation in MOOCs and Online Discussions

Irrespective of differences in their pedagogies, all MOOCs appear to face the common problems of high attrition rates. Large proportions of learners who enrol on courses never participate and many others leave courses after their first visit. The 2012 study by Rodriguez identified a dropout rate of 85% in Stanford-AI courses and 40% in connectivist MOOCs [3]. This trend of decreasing participation rates, associated with low retention in MOOCs, is described by Clow as *the funnel of participation* [10], and appears to show similarities with the previously observed participation ratio in online discussion forums. Studies of

2. https://www.udacity.com/us

discussions have demonstrated that no matter how high their volume, interactions are usually dominated by a small number of people, who post the largest amount of comments [11], [12]. Even in connectivist MOOCs, it has been reported that the 78% of the collaborative content was created by only 21% of its participants [10].

This is akin to the 90:9:1 Principle (van Mierlo's 1% Rule) which observed inequity online systems supporting behaviour change in that 90% of participants are passive *Lurkers*, 9% *Contributors* who contribute sparingly, and 1% *Superusers* who create the vast majority of the content [13]. We propose that identifying and encouraging *lurkers* in MOOCs to actively participate in their learning process might be used to boost learners' engagement.

1.2 Completion in MOOCs

There is no standard metric for measuring the completion and participation in MOOCs, so it can be challenging to compare different MOOCs. FutureLearn CEO, Simon Nelson, in a blog post[3], used data from FutureLearn alongside data shared by other MOOC providers. He reported completion rate of 8% at Harvard and MIT compared to 12% in FutureLearn.

Although some MOOCs have slightly better rates, completion of the majority of the course content remains low, prompting discussions around the possible reasons, impacts, and interpretations of this high attrition rate in MOOCs [14], [15]. There is evidence that some MOOC learners join a course only to follow one specific lecture or simply to have a MOOC experience. Koller et al. suggest, if learners leave the course before it is finished, their leaving early should not be considered as a failure or a loss to the learner, as long as their expectations have been met [15]. On the other hand, there is some evidence that many learners leave courses even though they initially had an intention of completing [14]. In their study, Khalil and Ebner investigate the reasons behind these high attrition rates, identifying some of the factors as follows:

- lack of time
- loss of motivation
- feelings of isolation
- lack of interactivity in MOOCs
- insufficient background knowledge and skills to cope with what is being taught in MOOCs.

Studies have shown that i) course completers are more interested in engaging with the course content and ii) learners who engage in social discussion forums are less likely to leave the course [16], [17]. FutureLearn takes a social constructivist approach in order to provide an environment that enables participants to easily reflect their opinion and interact with others for better social engagement. To achieve this, the platform inserts features facilitating social communication throughout the course adopting a Twitter-like *follow* system to help track and sustain interactive communication. This paper presents a study which analyses the use of the *follow* feature and its relation to the completion status of the people who use it.

Given the typically large numbers of MOOC participants, diagnostic analytic tools can be a particularly valuable means to inform educators about their learners' progress. Among the strengths associated with learning analytics and educational data mining, Papamitsion and Economides identified the ability to reveal critical moments and patterns of learning and to gain insights into learning strategies and behaviours [18]. This information can then be applied to provide timely learning support and interventions.

In a preliminary study, we sought to understand how MOOC learners engage with the social discussion forums by examining the *Developing Your Research Project (DYRP)* MOOC from different angles [19]. That study examined learner interactions within and across its eight week duration. We found that i) learners who contributed to online discussions do not usually have interactions with other participants, ii) if they do, the interactions do not fall into any regular pattern, iii) those learners who frequently interacted completed nearly all steps in each week.

To follow that preliminary study we analysed the *follow* interactions and completion status of socially active learners who wrote at least one comment to discussion threads or who follow at least one learner from the same course.

3. https://about.futurelearn.com/blog/completion-rates/

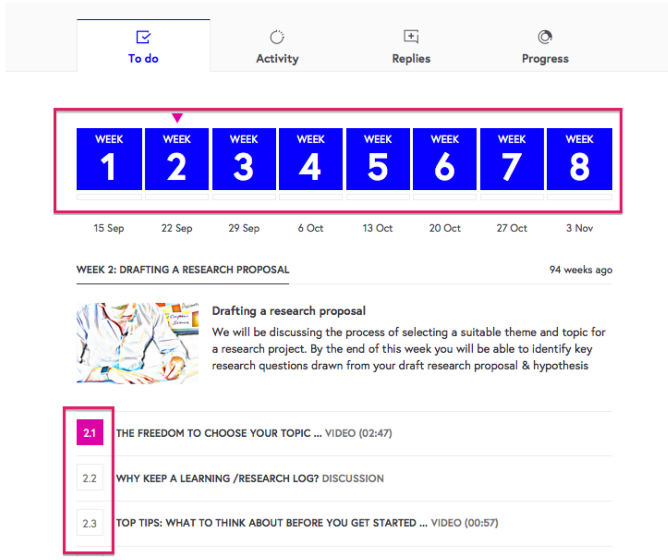


Fig. 2. FutureLearn MOOCs are structured around weeks and series of steps associated with weeks.

1.3 Objective of the Paper

Follow is one of the distinctive FutureLearn features that allows people to follow each other and directly access the contributions of those they follow on discussion boards. Participants are specifically able to see the comments posted by learners they have followed with a single click in the personal dashboard (see Fig. 1). This study investigated:

- behaviours of learners who follow someone from their course (see Section 2)
- the pattern of *follower* learners' participation in the discussion forum
- *follower* learners' completion of the course.

Since the follow feature is not commonly used in MOOC platforms, the analysis of *followers'* interactions has not yet been deeply investigated. The paper aims to address the following questions:

- How frequently do learners follow someone during a course?
- Do learners who follow someone make a greater number of active contribution to discussion forum?
- Do learners who follow someone in a course complete the course?
- How can we characterise the differences between completion rates comparing follow and discussion contribution behaviours?

Section 2 further identifies the social feature of FutureLearn platform and the available datasets we have analysed. Section 3 presents findings to answer the research questions presented above. Section 4 presents a discussion of the state of the art in prediction models in MOOCs. Section 5 discusses the conclusion and the findings, including implications for developing a predictive model based on learners' social presence. It also identifies opportunities for possible future research directions.

2 Social Features of FutureLearn Platform

Table 1 summarises the functionality and features investigated in this study.

FutureLearn MOOCs are structured in weeks and the series of steps associated with each week. The recommended route for learners to study is a logical progression in steps, but it is not compulsory. Learners manually mark each *completed* step as they progress. Additionally, FutureLearn has its own design prompting online discussions [8]. Each step in a week has an associated discussion board, designed as Twitter-like threads which enable the learners to scroll down and read sequentially through the set of associated comments. A learner can *like* a comment and *reply* to any specific comment. Additionally, learners are able to *follow* other participants in the platform by using the *follow* button. When they click on

TABLE 1
Specific functionality and features in the FutureLearn MOOC platform

ROLES
educator: Course designers or mentors
learner: Participants not from the educator team
ACTIONS
follow: Action of following someone in order to be informed of comments posted by that specific learner
post: Action of posting a comment to a thread
reply: Action of replying a comment in a thread
STATUS
follower: Follows other participants
followee: Is followed by a participant
poster: Posts to discussion threads
replier: Replies a comment
course (overall) completer: Completes at least half of the all steps of the final week of the course. If not, classified as **course non-completer**. We chose 50% as reference since FutureLearn describe *fully participated learners* as so (Test completion conditions are not considered in this research).
week completer: Completes at least half of the steps in a week. If not, classified as **week non-completer**.

a person's name on the discussion board, the person's profile is opened with the option of *Follow*. Learners are able to check the comments only posted by the people they follow (Fig. 1).

We analysed the University of Southampton's *Developing Your Research Project* MOOC, which ran from the 15th September - 9th November 2014. The datasets provided by FutureLearn are a snapshot of the participants' activities observed from the 15th September - 22nd November 2014. Anonymised data from every learner is stored by FutureLearn. The source data which we analysed was a subset drawn from the standard datasets: *enrolments, end of course, step activity*, and *comments*. FutureLearn also provided a static *followings* dataset upon our request. This dataset contained *follow* interactions amongst participants between the first day of FutureLearn and 2015-09-16 09:45:43 UTC), tracking around 1.2 million following relationships in the platform. We examined those data associated with the instance of the DYRP course selected for this study (2927 items). Table 2 summarises the types and attributions of the datasets. An algorithm was developed and written in Python in order to analyse the data; Matlab, GLE, and Power Point tools were utilised for the data visualisation.

3 Analysis and Results

3.1 General Statistics on Followers

Fig. 3 summarises the funnel of participation in the DYRP course. After the course was announced, 9855 learners enrolled, 5086 (51.6%) participants actually visited the course pages after the course started. Of these, 3852 (39%) completed at least one step and 2631 (26.7%) revisited the course and completed further steps. In total 1867 (18.9%) of 5086 learners participated in online discussions by writing at least one comment. Of these 789 participants followed discussions of one or more other course participants i.e. approximately 8% of all enrolled learners or 16% of the learners who actually visited at least once.

Fig. 4 and Fig. 5 illustrate that the volume of follow interactions accompanies a decline in resources accessed and weekly progress, occurring alongside the previously identified decline in discussion

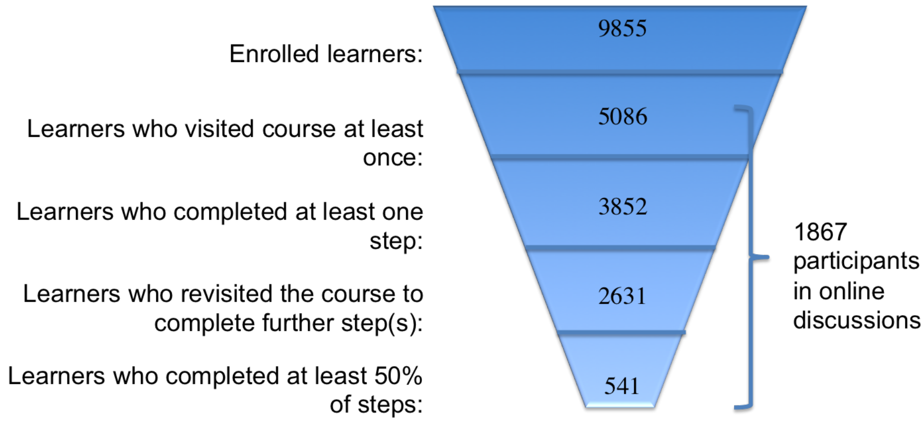


Fig. 3. Funnel of participation as observed in DYRP MOOC [19].

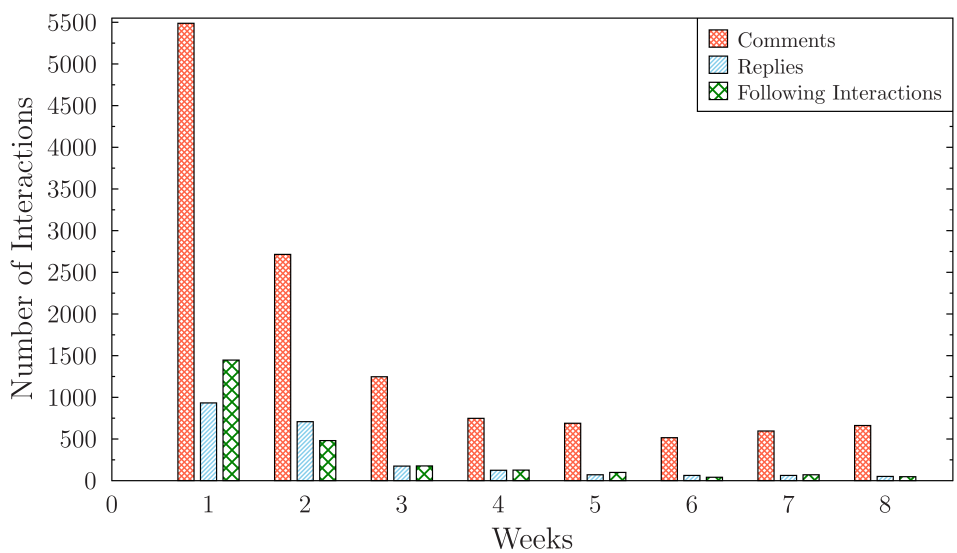


Fig. 4. Volume of weekly social activities: comments, replies, and followings

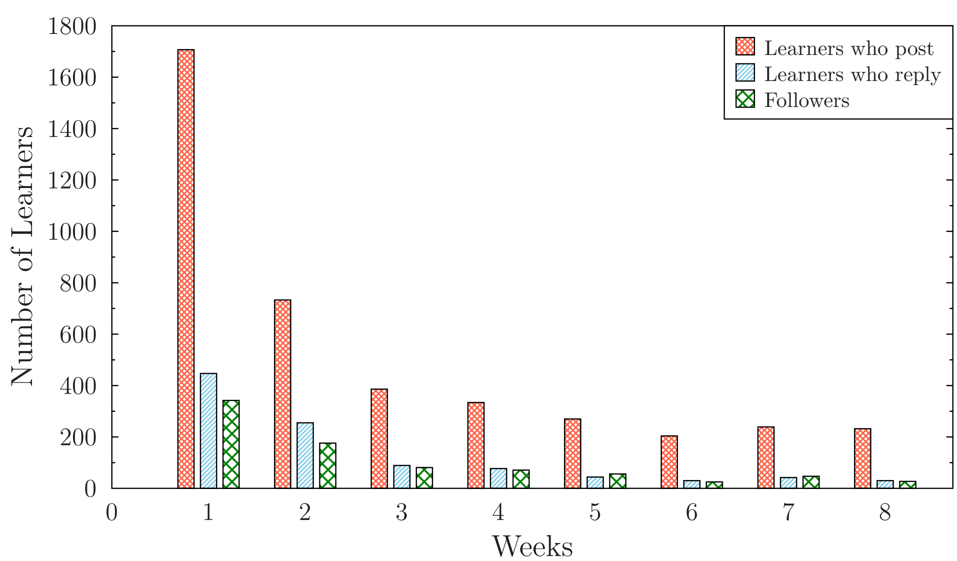


Fig. 5. Volume of weekly participants who are either a Commenter, Replier, or Follower

TABLE 2
List of FutureLearn Datasets and their Attributes

End of Course Stats
Overall participation rates in a MOOC i.e. number of those enrolled in the course and those who left the course
Enrolments
Enrolment records of participant Attributes: *learner_id, enrolled_at, unenrolled_at*
Step Activity
Number of steps completed by learners i.e. those checked the "completed" mark Attributes: *learner_id, step, week_number, step_number, first_visited_at, last_completed_at*
Comments
Records on the forum activities. This dataset identifies who posted: whether it was a reply, post timestamp, content and number of likes received. Attributes: *id, author_id, parent_id,step_ text, timestamp, likes*
Followings
Records on *follow* relationships amongst participants Attributes: *followed_user_id, follower_user_role, follower_user_id, follower_user_role, created_at*

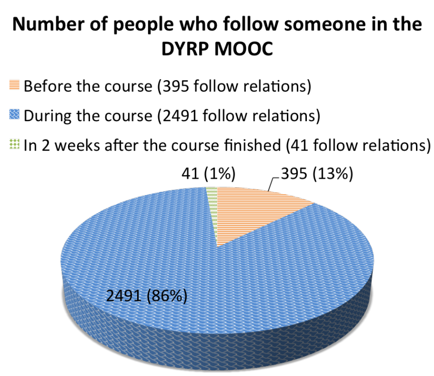


Fig. 6. Learners according to the time they start following somebody.

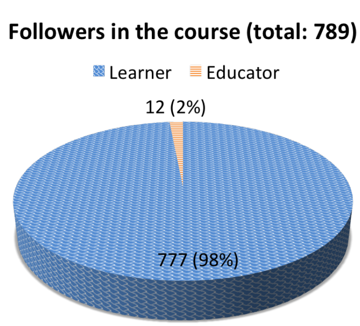


Fig. 7. Learners and educators who initialised follow interaction.

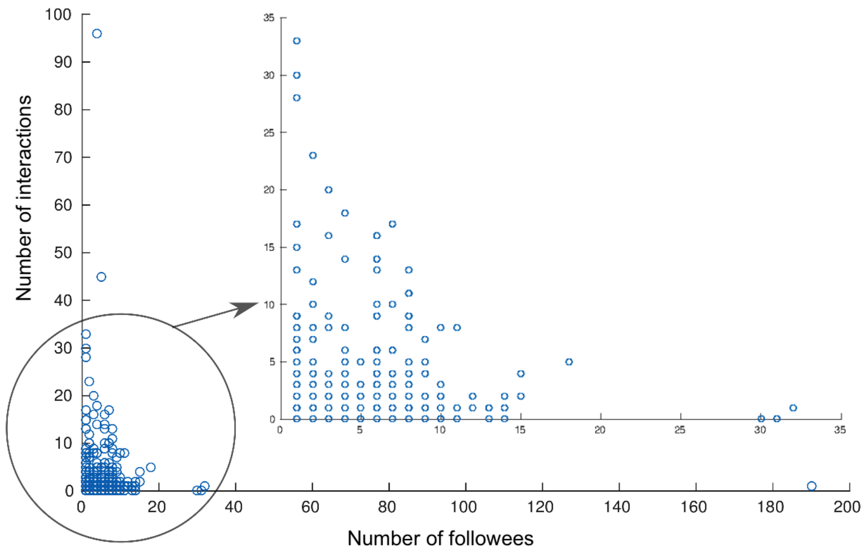


Fig. 8. Comparing the number of people whom a learner follows and the number of people with whom a learner interacted in the discussion forum.

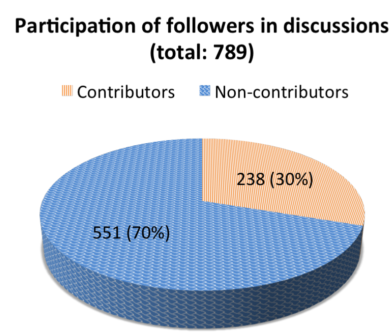


Fig. 9. Proportion of those who followed who contributed to discussions.

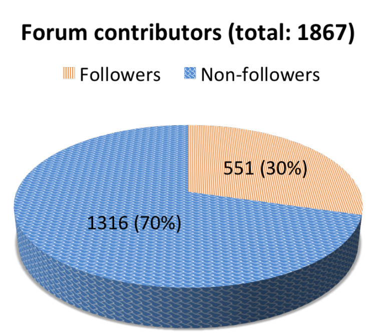


Fig. 10. Comparing discussion contributions between whose who did or did not follow others.

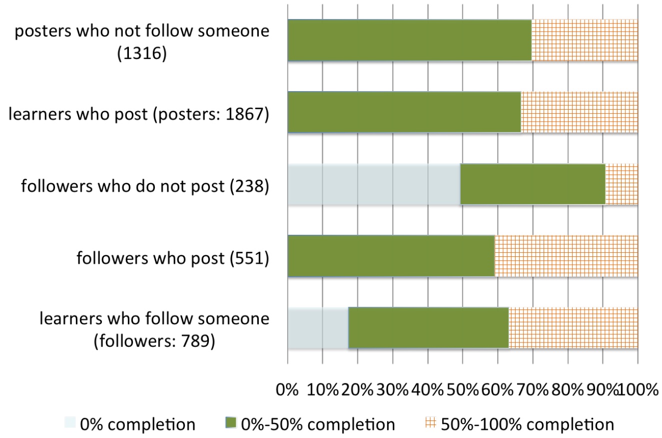


Fig. 11. Proportions of completers and non-completers of learners in different categories.

contributions [19]. Fig. 4 shows the number of activities initiated over the eight weeks. Fig. 5 illustrates the number of people who initiated those activities. It shows a weekly breakdown of the follow interactions of the 789 learners and educators whose distribution is shown in Fig. 7. The largest volume of follow interactions occurred in Week 1, it had the largest number of i) participants who completed course activities; and ii) comments posted to the discussion forums.

Fig. 6 shows the distribution of learners according to circumstances when they began following someone (before the course, during or after the course concluded). Since some learners had previously participated in a FutureLearn course(s), these learners may have already followed some individuals who also went on to participate in the DYRP MOOC. Over a hundred of such participants already had a follow relation when they enrolled in the DYRP MOOC.

It could be reasonable to assume that since these participants had already taken another MOOC together and had then subsequently enrolled together on the DYRP MOOC, they might be more likely to interact with each other. Nevertheless, our investigation shows that none of these prior experienced FutureLearn MOOC participants ever interacted with each other during the DYRP course. Indeed, interacting with each other in a previous MOOC, showing interest and enrolling in the same course again does not guarantee that these learners would be interested in each other's comments one more time. Additionally, we observe a small number of learners who joined the course late, and started following other learners shortly after the official end date of the course (Fig. 6).

Fig. 7 shows by role, the number and percentage of those participants who followed somebody. The majority of followers are not course educators. Note that the term *educator* is used for referring to course designers and mentors. Mentors are the experts from the ground i.e. PhD students are responsible for monitoring discussions during a MOOC. Leon et al. examine how mentors intervene during discussions on FutureLearn [20]. They observe that discussions often centre around a small number of people, most often educators and course facilitators. However, the *following* relations initiated between a follower and a followee are more widely distributed.

3.2 Involvement of Followers in Discussions

Fig. 8 to 10 examine the contributions to discussions in relation to whether the participants chose to follow others. Fig. 8 provides an overview of the size of each individual's network in discussion forums and the number of people that they follow. The majority (70%) of those participants who followed at least one other person contributed to the discussions. At the same time there was a small number of participants who commented extensively but who did not follow any other participants, 70% of all forum contributions were generated by those participants who followed no one.

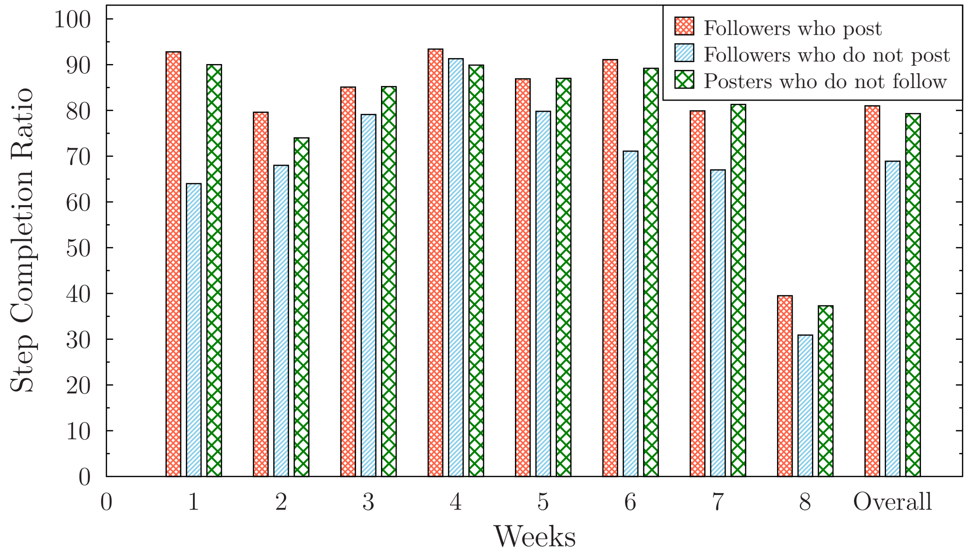


Fig. 12. Average percentages of the completed steps by learners in different categories.

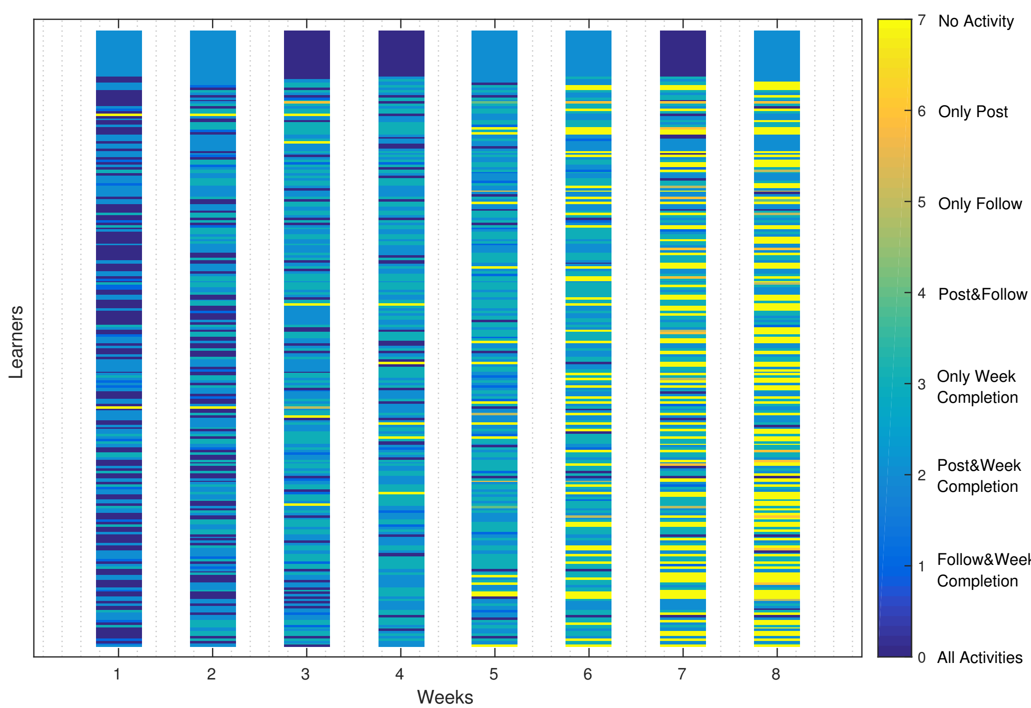


Fig. 13. Activities of completer learners [who followed at least once] throughout the course week-by-week (key on right).

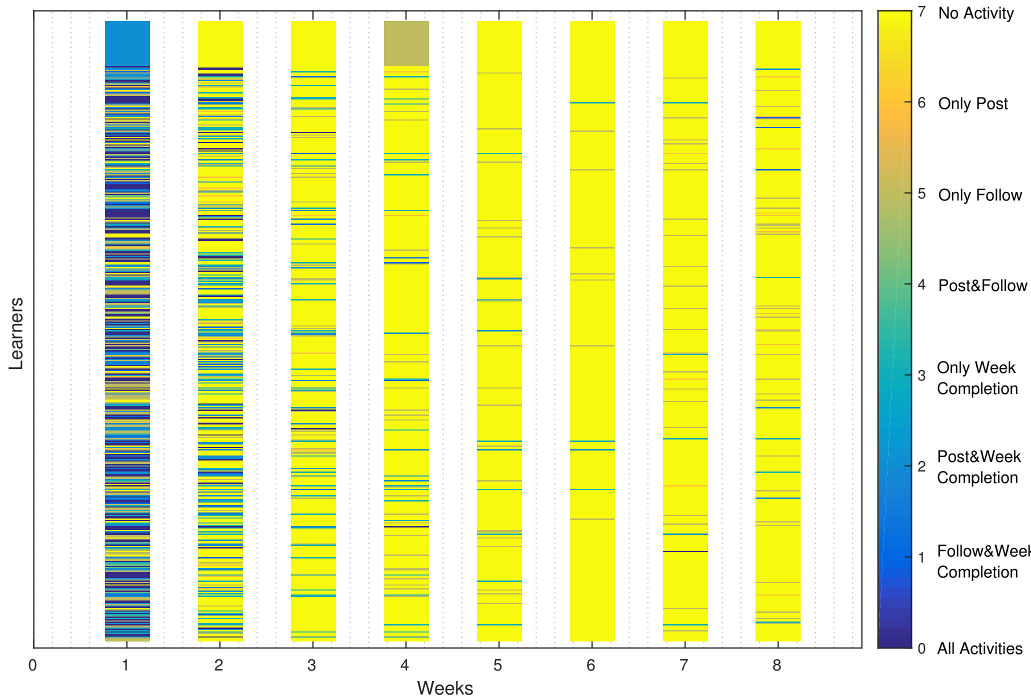


Fig. 14. Activities of non-completer learners [who followed at least once] throughout the course week-by-week (key on right).

3.3 Completion Success of Followers

The next step in the analysis was to examine course completion success amongst the followers. The DYRP MOOC is composed of 80 steps spread across eight separate weeks (Table 3). The number of weekly steps varies. For example, while Week 8 has 13 steps, Week 4 has only 6 steps. In order to provide a consistent representation, the proportion of steps in each week is analysed rather than the actual number of steps. Learners who completed at least 50% of the steps in a week are considered as a *completer* of the week; otherwise, the learner is named as a *non-completer* of the week.

TABLE 3
The number of steps in each week.

Weeks	1	2	3	4	5	6	7	8	Total
Steps	11	12	12	6	7	8	11	13	80

Fig. 11 shows the proportions of completer and non-completers by week, categorised according to their social activities. Learners are allocated across five categories, which are: i) learners who follow (aka *follower*), ii) followers who contribute to the discussions, iii) followers who do not contribute to discussions (aka *lurker*), iv) learners who contribute to discussions by posting (aka *poster*), v) and posters who do not follow.

One important observation is that every learner who posted to a discussion thread completed at least one step in the course. As shown in Fig. 11, if a learner is socially passive, it is likely that they will complete none of the steps, i.e. over 40% of socially passive followers did not complete any of the steps. The proportion of course completers is high if learners are socially active. Moreover, a larger proportion of course completers (41%) is observed amongst the learners who follow and post. The learners who either post or follow make up a similar percentage, slightly over 30%.

Fig. 12 plots the ratio of completed steps on a week-by-week basis for the learners. Learners are categorised by three distinct behaviours: i) those followers who post to discussions, ii) those followers who do not post to discussions, and iii) those posters who do not follow. We observe:

- Learners in each of the categories, regardless of whether or not they are an overall completer, progressed through the individual weekly steps at different rates.
- Fairly high weekly completion rates [from 60% up to 95%] are observed for all learners in each category throughout the course.
- The only exception is in the last week where the average fell to slightly over 30%.
- Followers who contributed to discussion threads completed the highest number of steps and represent the largest proportion of overall completers (Fig. 11).
- Posters who did not follow anyone completed more of steps than the followers who were socially passive in discussions.

These findings confirm those of our previous study which suggests that learners who repeatedly interacted with their fellows and actively contributed to discussions completed nearly all the 80 steps [19]. Another finding of our previous study is that the average step completion in DYRP is 26 of 80 possible steps (slightly over the 30% of the steps) [19]. However, *followers* who did not contribute to the discussion forum also performed better than the course average in completed steps (Fig. 12), implying that follow behaviours of learners could be used as an indicator for predicting their course completion.

Fig. 13 and Fig. 14 trace the activities on a week-by-week basis of every participant who was a follower (789 learners). Possible activities include completing a week, following, contributing to discussions, or combination of these activities (see Table 1).

These activities are shown with the aid of colour code, which has been chosen to remain readable when rendered or printed in black and white. Yellow (code 7, lightest) represents no activity; learners who neither participated in discussions nor followed anyone and did not complete more than 50% of the steps. Orange (code 6) and lime green (code 5) show learners either contributed to discussions or followed someone, but did not complete the week. Green (code 4) represents socially active non-completer learners. Turquoise (code 3) shows socially passive completer learners. Light blue (code 2) shows the learners who completed the week and were active in the discussions. Blue (code 1) represents learners who completed the week and followed someone. And finally, dark blue (code 0, darkest) shows learners who initiated all the possible activities. In a nutshell, the darker the colour, the more intense the learners' participation.

Fig. 13 shows the activities of completing followers while Fig. 14 shows non-completing followers. Although there are some similar behaviours amongst learners, the predominant activity profile for completers and non-completers are distinctive.

Course Completers: The social activeness of the course completers were sustained until Week 6. After Week 6, they showed limited activity. They hardly posted or followed other participants or completed the week. Full participation based on three behaviours (post, follow, step completion) was most prevalent in Week 1.

Course Non-completers: They have also been the most active in Week 1. Their level of activity and weekly completion declined sharply in Weeks 2 and 3 i.e. this is much earlier than course completers. Although no activity was observed in common especially after Week 3, it is still seen that a few of the non-completers kept on following someone or contributing to the discussions or very rarely completing the weeks. It appears that their behaviours are in accordance with the behaviours of lurkers in general discussion forums [13]. This guides us to think that the learners who read the comments and followed other participants or only concentrate on completing the steps become lurkers as the 90:9:1 principle proposes.

4 Related Work

One way in which the observed behaviours of learners in MOOCs discussed above might be used more generally is as part of a predictive model. For example, learners' social behaviours and completion rates could be used as parameters to predict whether or not a learner will socially participate in a coming week or complete the course at the end.

Predictive models are commonly used for making decisions by forecasting outcomes with the aid of statistical models and machine learning [21]. Predictive models are applied in various contexts ranging from health and politics to business and education. In MOOCs, predicting future participations and dropouts could be useful for detecting the need for educational interventions and the appropriate timing of such interventions. A number of researchers have attempted to apply predictive models. Focus includes anticipating learners' behaviours and identifying learners at risk. Tables 4 and 5 summarise available state-of-the-art techniques considering their objectives and their prediction methods. Table 5 summarises the dropout definitions identified in the literature and their notable findings.

There is no formal dropout definition, each study implements their own experiments using a variety of definitions. Two most widely-used definitions for dropout are:

1) **Not completed the final week:** If a learner does not engage in the final week's activities, they are assumed that they dropped out of the course. A similar assumption, proposed by some researchers, is that learners are marked as dropped out if they did not submit the final assignments.
2) **No activity during the most recent week** or **No further activities in the following weeks:** This definition differs from the previous in terms of the timing of the dropout. For example, if a learner's last activity is recorded in the fourth week of a six-week-long MOOC, that student is marked as "dropped out in the 4th week".

Several kinds of data are used and collected throughout the duration of a MOOC in order to observe learners' behaviour and develop prediction models. Typically four types of dataset are available: i) pre- and post- course surveys, ii) clickstream, iii) the results of assignments, and iv) activities in discussion forums. Some researchers use only clickstream data i.e. Amnueypornsakul et al. [30] and Kloft et al. [32], others combine the use of clickstream data, assignments and forum data. The studies were examined and selected to identify the strongest indicators that would have the most impact on prediction of dropouts. Factors which suggest strong correlation to dropout are listed below.

- Learners who show even minimal interaction in the forum after Week 1 are unlikely to drop out [22].
- Learners who start a course earlier and contribute to discussion forums are less likely dropout than others [23].
- Learners who lost their close peers are less likely to continue participating in the course forum [24].
- Learners who join later and participate in the least number of activities drop out [25].
- Assignment submissions are the most predictive [26].
- The length of forum posts is more strongly predictive than the number of posts and responses [26].
- Social integration in Week 1 is strongly correlated with course completion [29].
- Attrition rates and learners sentiment towards assignments and course materials are correlated [33].
- Learners' self-statements about their intention are more strongly predictive than demographics [36].

As presented in Tables 4 and 5, many studies consider learners' social participation as a factor for predicting dropouts i.e. [23], [24], [26], [29], [33]. For example, Chaplot et al. [33] use learners' sentiments extracted from the posts in the forum, while Jiang et al. [29] takes into consideration learners' level of activity in the forum in the first week.

We also consider building a prediction model using only social participation of learners. The analyses presented in this paper indicate that learners' *follow* behaviour along with their contribution to discussions may have a potential to predict learners' participation in a coming week.

5 Discussion and Conclusion

Researchers have combined different types of datasets, such as click-stream data, course surveys, assignment performances, and discussion forum activities, to have greater insight into learners' behaviour and success [37], [38]. They have analysed the relationships between learners' behaviours in MOOCs and their course completion rates to

- identify possible reasons for low retention rates [14],
- provide necessary help to learners [39],
- predict learners' future behaviours before they initiate [37].

TABLE 4
State-of-the-art of techniques for predicting learners' participation in MOOCs.

Study	Focus	Prediction Model	Datasets
Balakrishnan & Cootzee, 2013 [22]	1) Predicting course attrition, 2) Patterns of learners' behaviours	Hidden Markov Model	Clickstream, Assessment performance, Forum activity
Yang et al., 2013 [23]	Social factors on dropouts	Survival Model	Forum activity
Yang et al., 2014 [24]	Peer influence on learners' retention	Survival Model	Forum activity
Sinha et al., 2014 [25]	1) Learners' activities' patterns, 2) Predicting course attrition	Baseline Ngram Model, Graph Model	Clickstream, Forum activity
Taylor et al., 2014 [26]	Predicting course attrition	Logistic Regression	Clickstream, Assessment performance, Forum activity, Wiki revisions
Halawa et al., 2014 [27]	Students at risk of dropout	A prediction model developed.	Clickstream Assessment performance
Ramesh et al., 2014 [28]	Predicting learners' survival	Probabilistic Soft Logic	Clickstream, Assignment performance, Forum activity
Jiang et al., 2014 [29]	Predicting earning certificates	Logistic Regression	Assessment performance, Forum activity
Amnueypornsakul et al., 2014 [30]	Predicting learners' retention in a given week	Support Vector Machine	Clickstream
Sharkey and Sanders, 2014 [31]	Predicting course attrition	Random Forest Model	Clickstream
Kloft et al., 2014 [32]	Predicting course attrition	Fisher Scoring, Support Vector Machine	Clickstream
Chaplot et al., 2015 [33]	Predicting course attrition	Artificial Neural Network	Clickstream, Forum activity
Mi and Yeung, 2015 [34]	Predicting course attrition	Recurrent Neural Network	Clickstream, Forum activity
He et al., 2015 [35]	Students at risk of dropout	Logistic regression	Clickstream, Assessment performance
Robinson et al., 2016 [36]	1) Predicting learners' success before a course starts, 2) Intention to earn certificate	Natural Language Processing	Pre-course self-assessment

TABLE 5
Milestones of dropout definitions used and remarkable findings of these studies.

Study	Dropout Definition	Findings
Balakrishnan & Cootzee, 2013 [22]	No activity during the most recent week	Learners who rarely/never check their progress page leave the course earlier. Those who show even minimal interaction in the forum after the first week are unlikely to drop early.
Yang et al., 2013 [23]	Not completed the final week	Learners who start the course earlier and contribute to discussions are less likely to dropout the course than others who do not.
Yang et al., 2014 [24]	No activity during the most recent week	Learners who lost their close peers are not likely to continue participating in discussion forums.
Sinha et al., 2014 [25]	No activity during the most recent week	Recency and frequency of learners' activities would be used to predict learners' pathway. Learners who join courses later and do not participate in many activities usually leave courses.
Taylor et al., 2014 [26]	No further assignment or assignment submission	The most recent four weeks are predictive. Submitting assignments is the most predictive. The length of posts is more predictive than the number of posts and responses in the forum.
Halawa et al., 2014 [27]	1. Absence for a period exceeding one month 2. View fewer than 50% of videos	Dropout is strongly related to one type of bad persistence pattern i.e. learners who are absent 14 days or more are red-flagged.
Ramesh et al., 2014 [28]	No activity during the most recent week	The middle phase of a course is the most important phase to monitor students' activity for prediction of dropout.
Jiang et al., 2014 [29]	Not completed the final week	Social integration with a learning community in Week 1 is strongly correlated to completion.
Amnueypornsakul et al., 2014 [30]	No activity during the most recent	Features related to quiz attempt and submission are reasonable predictors in a given week.
Sharkey and Sander, 2014 [31]	No activity during the most recent week	Extracted 15 different data features related to learners' engagement and activity are strong predictors for dropout.
Kloft et al., 2014 [32]	No activity during the most recent week	Predictions are better measured at the end of a course.
Chaplot et al., 2015 [33]	No activity during the most recent week	There is a correlation between attrition and attitude of learners towards course materials

Since different MOOC platforms take different pedagogies and offer distinctive technological affordances, researchers use a range of parameters of MOOC learners' online behaviours to predict behaviours. The FutureLearn platform, which is used in our study, takes a social-constructivist conversational framework model for designing MOOCs as explained in Section 1 and 2. Having reviewed the literature, we are proposing a new approach. To the best of our knowledge, no researchers are currently using learners' forum interactions combined with their follow behaviours to predict possible dropouts. Therefore, we focused our attention on the recurrent interactions of the learners and their followers on the platform in order to estimate the possible dropouts.

Our previous study applied social network analysis techniques to investigate the correlation between learners' continuous interactions in discussions and their completion rates [19]. That study suggested that, for a very small number of learners who have continuous friendship with their peers in discussions, their completion rate is almost 95%.

In this study, we examined another distinct social feature of FutureLearn: the *follow* function. We believe in the value of this feature because it could provide us with greater insight into social behaviours of learners in their learning networks. Our research has focused on

- discerning the distinctive features of *follow* behaviours of MOOC learners,
- analysing *follower* learners' completion rates,
- identifying the possible value of *follow* behaviours for predicting learners' future participation in MOOCs.

We therefore addressed the following research questions:

How frequently do learners follow someone during a course? According to our findings, only a small fraction of learners attempt to use the follow feature provided by the platform. Not all forum participants use the follow feature. 70% of forum participant do not follow anyone (Fig. 10). There could be several reasons behind this.

1) **Lack of awareness:** Course providers usually advertise the follow feature in the introduction page but do not promote it throughout the course. Therefore, some learners may possibly not be aware of the follow function.
2) **Usability of follow:** After a learner starts following someone and interacts with them, the learner must manually control updates. After our study was completed, FutureLearn introduced on-site notifications in January 2016[4], which may improve the use of the *follow* feature.
3) **Confusion of learners:** Learners may not be able to decide who to follow. Brinton and Chiang suggest there is a problem of learners sometimes struggling to find a study partner resulting in a lack of interaction in discussion forums [40]. Learners may need help in finding like-minded study partners or conversations to initiate supportive interactions in MOOCs. It is important, therefore, to link learners to the right partners and to the right information in the discussion forums. In order to improve learners' participation and interactions, researchers have been implementing personalised services [41]. A personalised support for finding the right person to follow could help learners and improve the potential of a follow feature.

Further qualitative study is needed to investigate these possibilities.

Do learners who follow someone make a greater number of active contributions to discussion forum? More than half of the followers actively joined the discussions by writing a comment or replying to somebody else's comment (Fig. 9). However, there is lack of evidence to say that following someone causes rates for increased contribution to discussions and higher completion of the course.

We have also analysed peer interactions that were initiated by followers. When a learner follows someone in a course by clicking on the follow button in their fellow learner's profile, it may be fair to assume that the learner has an intention to follow further updates from their fellow learner and maybe to interact with them. However, only a small number of interactions occurred between followers and their followee (Fig. 8). Possible reasons could be various, summarised below:

4. https://about.futurelearn.com/blog/on-site-notifications

- The followee (the learner who was followed) either was no longer active in discussions or left the course. Therefore, the follower cannot interact with them.
- The follower could *lurk*, reading the comments but not commenting on them. In this case, it is almost impossible to assess from data whether or not they benefit from their followee.
- The follower may drop out the course after they start following someone.

If these reasons can be identified from available datasets of learners' activities in a future study, then it might be understood why learners stop interacting with their peers. Having already discussed learners' struggles with finding peers to communicate with (and follow), a positive impact on participation in discussions might be achieved by matching learners with appropriate peers and encouraging them to follow more people.

Do learners who follow someone in a course complete the course? Over 30% of the followers completed the course by completing at least half of the steps. Fig. 13 and Fig. 14 demonstrate activities of the completers and non-completers throughout the MOOC. According to our findings, completers are usually more active participants in course activities and use available features in the platform.

This finding corroborates suggestions of previous studies that, socially active learners are the biggest portion of the course completers [16], [17], [19]. Indeed, some studies have used social behaviours of learners for developing a prediction model of course completion (Tables 4 and 5). Our findings indicates that learners' social behaviours including their follow interactions could also be used for predicting learners' future participation.

How can we characterise the differences between completion rates comparing follow and discussion contribution behaviours? We categorised active social participants as i) followers who contributed to discussions, ii) followers who did not contribute and iii) posters who did not follow anyone. Fig. 12 compares the percentage of their completions rates. While the followers who contributed to discussions completed over 80% of the steps in the course, the followers who did not contribute to discussions remained lower (60%). The course completion rate of those who only post is slightly smaller than the completion rates for followers who posted to discussions.

These findings also reiterate the relationship between presence in discussions and course completion. Some researchers propose that applying a predictive model to a personalised service might be a constructive means to boost learners' online participation. For example, Rose et al. [42] use learners' participation in discussion forums to predict attrition by analysing learners and their cohort, who are at similar place in the course. Given our findings future directions of MOOC research include

- developing personalised tools based on prediction models and social behaviours of learners,
- investigating the impact of those tools on participation in MOOCs and course completion.

Overall our research confirms previous findings that participation in course forums is a good indicator of committed participation in a course, and that learners who fully participate are the most likely to complete. Furthermore, it is clearly the case that if a learner follows another learner they are demonstrating that they are actively participating in the course, even if their participation does not extend to making original posts to the course forum. An original contribution of our work is to show that identifying such lurkers provides us with another useful parameter to feed into the model for predicting likeliness to complete.

ACKNOWLEDGMENTS

We would like to thank Dr. Halil Yetgin for the technical help he kindly offered during the analysis process.

REFERENCES

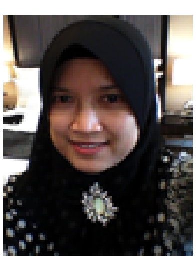
[1] G. Siemens, "Connectivism: A learning theory for the digital age," *International journal of instructional technology and distance learning*, vol. 2, no. 1, pp. 3–10, October 2005.
[2] A. Fini, "The technological dimension of a massive open online course: The case of the CCK08 course tools," *The International Review of Research in Open and Distributed Learning*, vol. 10, no. 5, November 2009.

[3] C. O. Rodriguez, "MOOCs and the AI-Stanford like courses: Two successful and distinct course formats for massive open online courses," *European Journal of Open, Distance and E-Learning*, vol. 15, no. 2, pp. 1–13, 2012.
[4] J. Kennedy, "Characteristics of massive open online courses (MOOCs): A research review, 2009-2012." *Journal of Interactive Online Learning*, vol. 13, no. 1, pp. 1–16, September 2014.
[5] M. Bali, "MOOC pedagogy: Gleaning good practice from existing MOOCs," *Journal of Online Learning and Teaching*, vol. 10, no. 1, pp. 44–56, March 2014.
[6] L. Guàrdia, M. Maina, and A. Sangrà, "MOOC design principles: A pedagogical approach from the learner's perspective," *eLearning Papers*, vol. 33, pp. 1–6, May 2013.
[7] J. Mackness, M. Waite, G. Roberts, and E. Lovegrove, "Learning in a small, task–oriented, connectivist MOOC: Pedagogical issues and implications for higher education," *The International Review Of Research In Open And Distributed Learning*, vol. 14, no. 4, September 2013.
[8] R. Ferguson and M. Sharples, "Innovative pedagogy at massive scale: teaching and learning in MOOCs," in *9th European Conference on Technology Enhanced Learning*. Graz, Austria: Springer, September 2014.
[9] D. Laurillard, *Rethinking university teaching: A conversational framework for the effective use of learning technologies*. Routledge, 2013.
[10] D. Clow, "MOOCs and the funnel of participation," in *3th International Conference on Learning Analytics and Knowledge*. Leuven, Belgium: ACM, April 2013.
[11] C. Yeager, B. Hurley-Dasgupta, and C. A. Bliss, "cMOOCs and global learning: An authentic alternative." *Journal of Asynchronous Learning Networks*, vol. 17, no. 2, pp. 133–147, July 2013.
[12] S. Jiang, S. M. Fitzhugh, and M. Warschauer, "Social positioning and performance in MOOCs," in *Workshop on Graph-Based Educational Data Mining, EDM 2014*, London, UK, July 2014, p. 14.
[13] T. van Mierlo, "The 1% rule in four digital health social networks: An observational study," *Journal of medical Internet research*, vol. 16, no. 2, p. e33, February 2014.
[14] H. Khalil and M. Ebner, "MOOCs completion rates and possible methods to improve retention-a literature review," in *EdMedia: World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Tampere, Finland, June 2014.
[15] D. Koller, A. Ng, C. Do, and Z. Chen, "Retention and intention in massive open online courses: In depth," *Educause Review*, vol. 48, no. 3, pp. 62–63, June 2013.
[16] Y. Wang and R. Baker, "Content or platform: Why do students complete MOOCs?" *MERLOT Journal of Online Learning and Teaching*, vol. 11, no. 1, pp. 17–30, May 2015.
[17] S. Joksimović, N. Dowell, O. Skrypnyk, V. Kovanović, D. Gašević, S. Dawson, and A. C. Graesser, "How do you connect?: Analysis of social capital accumulation in connectivist MOOCs," in *5th International Conference on Learning Analytics And Knowledge*, Poughkeepsie, New York, March 2015.
[18] Z. K. Papamitsiou and A. A. Economides, "Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence." *Educational Technology & Society*, vol. 17, no. 4, pp. 49–64, October 2014.
[19] A. S. Sunar, N. A. Abdullah, S. White, and H. C. Davis, "Analysing and predicting recurrent interactions among learners during online discussions in a MOOC," in *11th International Conference on Knowledge Management.*, Osaka, Japan, November 2015.
[20] M. León, S. White, S. White, and K. Dickens, "Mentoring at scale: MOOC mentor interventions towards a connected learning community," *EMOOCs 2015 European MOOC Stakeholders Summit*, pp. 13–17, May 2015.
[21] S. Finlay, *Predictive Analytics, Data Mining and Big Data: Myths, Misconceptions and Methods*. Palgrave Macmillan, 2014.
[22] G. Balakrishnan and D. Coetzee, "Predicting student retention in massive open online courses using Hidden Markov models," *Electrical Engineering and Computer Sciences University of California at Berkeley*, May 2013.
[23] D. Yang, T. Sinha, D. Adamson, and C. P. Rose, "Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses," in *NIPS Data-Driven Education Workshop*, Nevada, USA, December 2013.
[24] D. Yang, M. Wen, and C. Rose, "Peer influence on attrition in massively open online courses," in *Educational Data Mining 2014*, London, UK, July 2014.
[25] T. Sinha, N. Li, P. Jermann, and P. Dillenbourg, "Capturing "attrition intensifying" structural traits from didactic interaction sequences of MOOC learners," in *EMNLP Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses*, Doha, Qatar, September 2014.
[26] C. Taylor, K. Veeramachaneni, and U. M. O'Reilly, "Likely to stop? predicting stopout in massive open online courses," *arXiv preprint arXiv:1408.3382*, August 2014.
[27] S. Halawa, D. Greene, and J. Mitchell, "Dropout prediction in MOOCs using learner activity features," in *2th MOOC European Stakeholders Summit*, Lausanne, Switzerland, February 2014.
[28] A. Ramesh, D. Goldwasser, B. Huang, H. Daume III, and L. Getoor, "Learning latent engagement patterns of students in online courses," in *AAAI Conference on Artificial Intelligence*, North America, June 2014.
[29] S. Jiang, A. Williams, K. Schenke, M. Warschauer, and D. O'dowd, "Predicting MOOC performance with week 1 behavior," in *7th International Confrence on Educational Data Mining*, London, UK, July 2014.
[30] B. Amnueypornsakul, S. Bhat, and P. Chinprutthiwong, "Predicting attrition along the way: The UIUC model," in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, October 2014.
[31] M. Sharkey and R. Sanders, "A process for predicting MOOC attrition," in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, October 2014.
[32] M. Kloft, F. Stiehler, Z. Zheng, and N. Pinkwart, "Predicting MOOC dropout over weeks using machine learning methods," in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, October 2014.
[33] D. S. Chaplot, E. Rhim, and J. Kim, "Predicting student attrition in MOOCs using sentiment analysis and neural networks," in *Fourth Workshop on Intelligent Support for Learning in Groups*, Madrid, Spain, June 2015.

[34] F. Mi and D.-Y. Yeung, "Temporal models for predicting student dropout in massive open online courses," in *IEEE International Conference on Data Mining Workshop (ICDMW)*, Atlantic City, NJ, November 2015.
[35] J. He, J. Bailey, B. I. Rubinstein, and R. Zhang, "Identifying at-risk students in Massive Open Online Courses," in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, Texas, USA, January 2015.
[36] C. Robinson, M. Yeomans, J. Reich, C. Hulleman, and H. Gehlbach, "Forecasting student achievement in MOOCs with natural language processing," in *Sixth International Conference on Learning Analytics & Knowledge*, Edinburgh, UK, April 2016.
[37] A. M. Shahiri, W. Husain, and N. A. Rashid, "A review on predicting student's performance using data mining techniques," *Procedia Computer Science*, vol. 72, pp. 414–422, December 2015.
[38] A. Dutt, S. Aghabozrgi, M. A. B. Ismail, and H. Mahroeian, "Clustering algorithms applied in educational data mining," *International Journal of Information and Electronics Engineering*, vol. 5, no. 2, p. 112, March 2015.
[39] A. S. Sunar, N. A. Abdullah, S. White, and H. C. Davis, "Personalisation of MOOCs: The state of the art," in *7th International Conference on Computer Supported Education*, Lisbon, Portugal, May 2015.
[40] C. G. Brinton and M. Chiang, "Social learning networks: A brief survey," in *48th Annual Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, March 2014, pp. 1–6.
[41] E. O'Donnell, S. Lawless, M. Sharp, and V. P. Wade, "A review of personalised e-learning: Towards supporting learner diversity," *International Journal of Distance Education Technologies*, vol. 13, no. 1, pp. 22–47, 2015.
[42] C. P. Rosé, R. Carlson, D. Yang, M. Wen, L. Resnick, P. Goldman, and J. Sherer, "Social factors that contribute to attrition in MOOCs," in *First ACM conference on Learning@ scale conference*, Atlanta, USA, March 2014.

Ayse Saliha Sunar received her BSc in mathematics at Gazi University in Ankara, Turkey. She then took a one year professional teaching programme to teach mathematics to secondary school students. In 2023, she pursued an MSc at the Graduate School in Social and Information Science Department at Nagoya University, Japan. Her MSc thesis focussed on learning languages with the aid of intelligent tutoring systems. Currently, she is a PhD candidate in the Web and Internet Science Research Group (WAIS) at the University of Southampton, UK where she contributes to the research of the MOOC Observatory. Her research interests include big data, learning analytics, MOOCs, learning at scale, technology enhanced learning, predictive models and personalisation of learning.

Dr Su White is an associate professor in the Web and Internet Science Research Group (WAIS) at the University of Southampton. Su leads the MOOC Observatory at Southampton, whose team is developing dynamic analytical dashboards to analyse learning behaviours, initially in near to real time. Together, researchers in the team are studying the evolution of educational practices associated with MOOCs and Open Learning as they impact on organisations, teachers and instructors and learners, the team is also involved in the production and facilitation of MOOCs running on the FutureLearn platform. Su has worked on the development of innovative educational technology projects since the early 1990's. Her particular interest is in understanding the way in which curricula can emerge from educational artefacts and how massive datasets afforded by large scale interactions in learning platforms can transform learners' understanding of their own progress and educators' understanding of the learning process.



Dr Nor Aniza Abdullah is an Associate Professor at the University of Malaya, Malaysia. She has a master degree in interactive multimedia from the Westminster University, London. She obtained her PhD degree in Computer Science from the University of Southampton. Her research interest is in personalisation and recommender systems, personal learning environments (PLE), multimedia processing and retrieval technologies. She believes that a true sense of a personalised MOOC system would excite community of learners to continuously seek knowledge and skills for the benefit of their community and people at large.

Professor Hugh C Davis is a Professor of Learning Technologies in the Web and Internet Science Research Group (WAIS) at the University of Southampton. He is also a University Director of Education and was until recently the Director of the Institute for Learning Innovation and Development (ILIaD) leading cross-university collaboration between faculties and professional services to research and transform education. His current research interests are all concerned with how technologies can change our perception and experience of learning, and include the reach and impact of MOOCs, personal learning environments (PLEs), educational repositories, educational analytics and semantic applications in education. He has extensive experience of applying research outputs to create real change in educational practice.