Watanabe, T. and Seta, K. (Eds.) (2015). Proceedings of the 11th International Conference on Knowledge Management.

Analysing and Predicting Recurrent Interactions among Learners during Online Discussions in a MOOC

Ayse Saliha SUNAR[a*], Nor Aniza ABDULLAH[b], Su WHITE[a] & Hugh C DAVIS[a]

[a]*Electronics and Computer Science, University of Southampton, Southampton, United Kingdom*
[b]*System and Technology, University of Malaya, Kuala Lumpur, Malaysia*

*ass1a12@soton.ac.uk

Abstract: High attrition rates are one of the biggest concerns in MOOCs. One of the possible causes may be learners' lack of interactions and low levels of participations in MOOCs online discussions. Research to measure and predict recurrent interactions of learners in MOOCs online discussions has the potential to gain inside into the likely impact on the attrition rate. It is argued that personalisation in MOOCs has the potential to increase learners' interactions and associated factors to continuous friendships. In this paper, a detailed analysis has been carried out of learners' interactions within a MOOC. This paper investigates learners' interaction habits and their recurrent interactions throughout the entire duration of a MOOC's course, and consequently proposes a method to measure the interactions and predict possible interactions between peers. The findings denote that when a learner interacted with their peer, they most probably interact again in the following weeks. Moreover, our proposed prediction method also demonstrate promising results towards predicting future interactions between learners based on their previous relationships.

Keywords: Technology enhanced learning, connectivism, MOOCs, attrition, retention, social learning networks, personalisation, recommender system

1. Introduction

MOOCs can be seen as an enhancement of open education resources. However, early experience of MOOCs indicates that there are further enhancements which can be made for improving MOOC education. The work reported in this paper is concerned with a further intervention to enhance massive open education.

Open access to educational resources has been intensively discussed since the beginning of the 2000s. In 2001, Massachusetts Institute of Technology launched the MIT OpenCourseWare project to publish all course materials online. In September 2007, The Cape Town Open Education Declaration was signed by hundreds of people in the educational fields. This declaration points out the importance of free information sharing at a global level for educational purposes. In order to promote open access to digital resources, the Open Educational Resources (OER) movement has emerged and many universities made their courses available online and publicly published course contents and materials such as presentation slides and filmed classroom lectures (Yuan, MacNeill & Kraan 2008).

In 2005, Siemens (2005) has proposed connectivist learning theory for the digital age. According to the theory, learning is a continual process and occurs in a variety of ways and technology shapes how and where we learn. In the age of technology, learning is a process of connecting people to each other, to ideas, to knowledge, and to information resources. Moreover, technologies such as web and social media tools provide opportunities that people can efficiently find information and make meaningful connections in order to acquire useful knowledge. Subsequently, Siemens and Downes opened their *Connectivism and Connective Knowledge'08* course to public in 2008. Over 2000 learners participated in the course. Its structure was based on the connectivist learning theory so that social media platforms suitable for peer communications are provided along with video lectures and resources for learners. Peer communication and group discussions on social media platforms are encouraged during the lectures. This open and free online course becomes known as a MOOC (massive open online course). Subsequently numbers of universities and private initiatives have started launching their own MOOCs. Even though not all MOOCs are based on the philosophy of connectivism, all courses promote online discussions at least through the online discussion forums embedded in the course.

However, concerns have been expressed on the different aspects of MOOCs including educational quality (Margaryan, Bianco & Littlejohn 2015) and high attrition rate (Clow 2013). Clow (2013) has described the participation pattern of MOOC learners as a funnel of participation. Thus, only a fraction of learners of those who registered the course engages in course activities. A group of engaging learners make substantial process and a large number of learners drop off at each stage.

There are researches that examine learners' behaviour in MOOCs to try to identify what reasons might be behind attrition rates. For example, Khalil and Ebner (2014) point out that some of the factors that influence attrition rates could be lack of time, learners' motivation, feelings of isolation and the lack of interactivity in MOOCs, and insufficient background knowledge and skills. Some other researchers try to predict attrition rate based on those factors so that the predictions may lead to an intervention to improve retention rates. For example, Kloft, Stiehler, Zheng and Pinkwart (2014) predict attrition in a MOOC over the weeks with the aid of machine learning. The authors have used click stream data such as total time spent for viewing a material, quiz submission and number of re-listen during video plays. Stein and Allione (2014) have also proposed a method for estimating the probability of dropping out. According to the results of their proposed method, learners who engaged in the course from the first days and submitted either the quiz or a peer assessment exercise were less likely to leave the course. Interestingly, learners who engaged in the social forum but not submitted the assignments were more likely to leave the course.

Contrarily, Yang, Wen and Rose (2014) have analysed learners' interaction on a discussion forum. They have looked into learners' direct relations, threads that learners posted, participation patterns and so on. The results show that when relationships between learners decline, the attrition is on increase. Additionally, the study of Jiang, Warschauer, Williams, O'Dowd and Schenke (2014) also indicate as a result of their analysis that there is a strong correlation between learners' social engagement in a learning community and course completion.

A number of systems which enable personalisation through recommender systems that have been developed and evaluated by researchers keen to enhance participants' retention. For example, Zhuhadar and Butterfield (2014) have analysed learners' logs on several social media tools promoted by the course designer and developed a recommender system for encouraging learners to be active on more numbers of social media tools. Yang, Piergallini, Howley and Rose (2014) propose a recommendation system for easing learners' experiences in discussion forums. The authors deliver personalised forum thread recommendation based on threads' features and learners' interactions in threads. However, those researches did not focus on building and prompting continuing interactions among learners. They have directly focused on the use of tools and forums.

On the contrary, the ultimate goal of the work reported in this paper is to 1) encourage learners to contribute to online discussions and 2) dynamically develop and maintain their social learning networks across MOOCs environment. To accomplish this aim, it is crucial to understand pattern of learners' interactions and contributions to online discussions particularly the recurrent interactions between learners because it could infer a more long lasting interactions or continuous friendship in the online learning community. Therefore in this paper, learners' interaction habits are analysed and a prediction method of recurrent interactions between learners over eight weeks of learning a MOOC course is proposed. In order to predict learners' interactions, not only existing peer interactions and contributions, but also hidden relationships based on learners' mutual friendships are considered.

2. Methodology

Since the case study is conducted on the FutureLearn platform, the general feature of the platform is first examined in Subsection 2.1. Later on, the method of this study is examined in Subsection 2.2.

2.1. Features of FutureLearn

FutureLearn is a UK based MOOCs platform which launched in September 2013. Since then, over 150 free courses are delivered and 1,724,238 learners have joined FutureLearn. Each course composes of weeks containing several video and written lectures, additional resources, activities and peer/quiz assessments. Figure 1 illustrates the general view of a week on a MOOC's page. Learners are able to choose any step of any week to study. When they finish a step, they simply mark as complete. If a learner has studied all steps in a week and completed the assessment if there is any, it means the learner have completed 100% of the week. In order to facilitate learners' interactions, each step has discussion board embedded in the page as it is shown in Figure 2. Additionally, some courses encourage learners to get interacted on external social media tools such as Twitter, Google Plus, and blogs. However, this work only considers social interactions within the discussion boards.

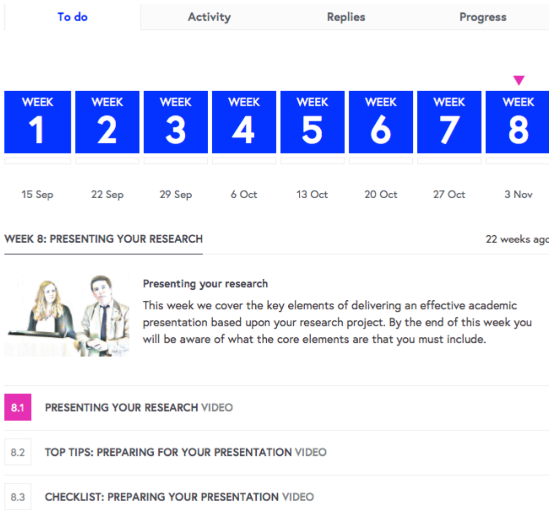


Figure 1. General view of a week in a MOOC on FutureLearn.

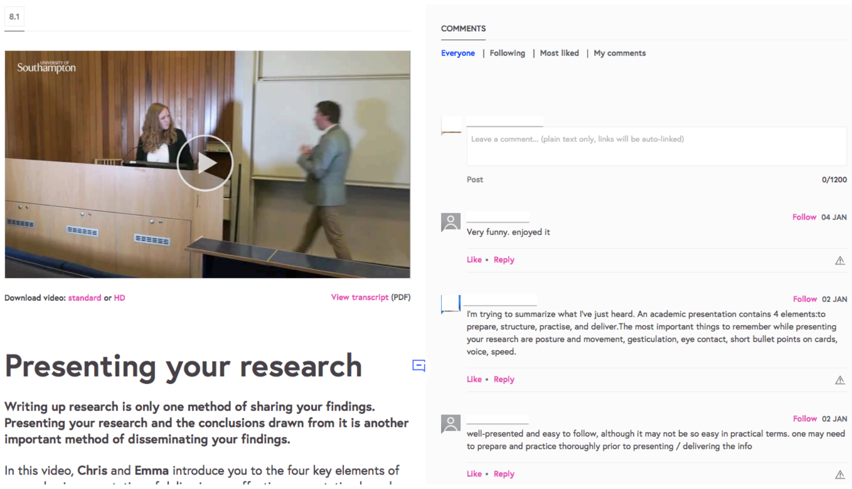


Figure 2. Discussion board on a step of a MOOC.

During the study, learners' contributions with comments are considered in two categories:

- *Individual comment*: posted single comments reflecting learner's opinion, thought, question and so on.
- *Interaction* (between two learners): reply to somebody's comment.

Social learning networks refer the connection between learners during their interactions on the discussion board. Thus, if a learner replies another learner's comment, this indicates an interaction from the learner (replier) to other learner (receiver), which is also called as friendship in this work.

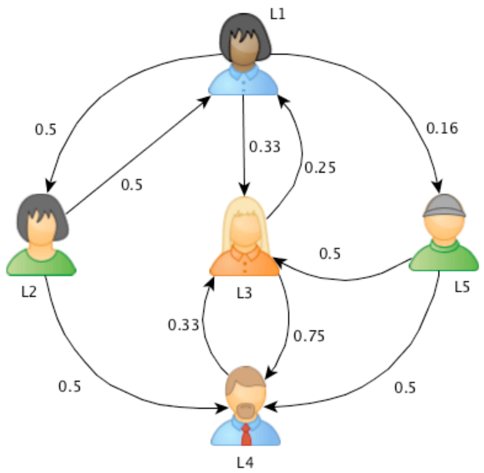
2.2. *Method of the study*

In this paper, we have made a case study in order to analyse learners' interactions, observe learners' tendency to repeatedly interact each other during online discussions, and test our prediction model. The MOOC for the case study is the *Developing Your Research Project* (DYRP) that was organised by the University of Southampton from the 15th September to the 5th November 2014. The basic analytical data is provided by FutureLearn, which includes overall simple analysis such as the number of enrolled people, the number of people joined online discussions and anonymised data of learners' activities (5th September-22nd November 2014) in online discussions. Anonymised data shows that who wrote a comment and who replied which comment. In order to further analyse this data, a utility is developed using Python programming language. The analysed data is visualised with Gnuplot graphing utility (see Section 4). We have used the developed utility for following steps:

1. Identify weekly data based on the number of contributions and interactions in online discussions
2. Identify strength of relationships between each learners in online discussions (see Section 3)
3. Predict possible relationships between learners in online discussions
4. Compare the actual interactions and predictions

3. Predicting Interactions

Online social networks have been analysed for several purposes such as product recommendation to users or friend recommendation in a social media tool.



Learner-Learner Interaction Matrix

	L1	L2	L3	L4	L5
L1	*	0.5	0.33	0	0.16
L2	0.5	*	0	0.5	0
L3	0	0.25	*	0.75	0
L4	0.33	0	0	*	0.66
L5	0	0	0.5	0.5	*

Figure 3. Graph and matrix representation of social networks.

In our research, it is important to understand how frequently a learner interacts with their peers in online discussion sessions. Therefore, the strength of relationships based on a peer's interactions is calculated. The method proposed by Yang, Guo, Liu, and Steck (2014) is adopted to represent the interactions and friendships among learners, and the strength of relationships (s) in a form of a graph $G = (L, F)$ and a matrix as shown in Figure 3. In the graph, L is the set of learners in a MOOC with $|L| = n$ and F is the set of friendship links. The friendship links are directed, which identifies the direction of relationships. For example, since learner L2 replies a comment of learner L4, the friendship links is directed from L2 to L4 in Figure 3. On the other hand, there is two-way relationship between L3 and L4. The strength of relationships in social learning network shown in the graph G is represented by a matrix $S \in \Re^{n \times n}$. The directed and weighted social learning relationship between learners u and v is represented by a strength value $S \in [0,1]$. If those learners never interact, then $S_{u,v} = 0$. The higher value represents frequent interactions between learners.

In this research, if learner u has interacted with learner v (by replying a comment of the learner v) during online discussions in a MOOC, then the strength of friendship (s) between the learner u and the learner v is estimated as follows:

$$s_{u,v} = \frac{t_{A_{u \to v}}}{t_{A_u}},$$

where $t_{A_{u \to v}}$ is the number of interaction from the learner u to the learner v and t_{A_u} is the total number of contributions the learner has done in the MOOC (refers to a certain week in a MOOC course). Since this formula calculates interactions between two specific learners, only the frequency of interactions between them is considered.

Table 1: Possible categories for prediction

Case 1: friendship with *zero-comment* learners Learners in this category have not contributed to the online discussions yet. Therefore, they have no social learning network and learning history in the MOOC. Thus, strength of a possible friendship cannot be predicted.	Cannot be predicted
Cases 2 – *persistent friendship*: friendship between learners who have been friends before The proposed system uses arithmetic to predict the strength of friendship between learner and learner v whom learner u has previously interacted with, as represented in the equation where n is the number of weeks taken by the learners u and v.	$s'_{u,v} = \sum_{i=1}^{n} \frac{(s_{u,v})_i}{n}$
Case 3 – *indirect friendship*: friendship with learner v through mutual friend(s) Even though learners u and v not directly interacted, they could have common friend(s) in some cases. The proposed system uses correlation between learners u and v through the mutual friend(s) j with the equation where k is the number of mutual friends of the learners u and v.	$s'_{u,v} = \sum_{j=1}^{k} \frac{s_{u,j} \times s_{j,v}}{k}$
Case 4 – *isolated friendship*: friendship with learner v who has no mutual friend(s) The proposed system use a probabilistic model for prediction of the strength of possible friendship in the case that the learners u and v share no mutual friends and have never interacted before. In this case, the only information the proposed system has is the previous activities of learners u and v. However, Case 4 is different from Case 1. While learners in Case 1 has no contribution to online discussions, learners in Case 4 have contributed to online discussions and likely have their own social learning network. However, their social learning networks have no intersection. In this case, learner's overall interest through past contribution to online discussions could be useful for predicting learners' possible interaction in the future. *Overall interest* is the social interest that a learner has showed from the beginning of the course until a current week. It is calculated as follows: $p_u = \frac{c_u}{c}$, where c_u is the total number of comments made by learner u, and c is the total number of comments made by all. Therefore, the predicted strength of friendship between the learners is calculated with the equation, where p_u and p_v are the *overall interest* through the completed weeks shown by both learners u and v, respectively.	$s'_{u,v} = p_u \times p_v$

If learner *u* has social learning network(s) in the completed courses, then the proposed method would predict the possible contribution of his social learning network to the current online discussions, and the strength of his possible friendships with other learners in a coming week of a MOOC's course. Details of the prediction method are presented in the following subsections.

3.1. Predicted Social Learning Network

In order to identify a learner's predicted social learning networks, predicted strength of friendships with every other learner needs to be first determined. However, the predicted strength of friendship between two learners varies according to their kind of friendship history. To categorise this relationship, our proposed method suggests four possible cases as explained in Table 1.

4. Experiments and results

4.1. General Analysis

Funnel participation (Clow 2013) has been observed in the studied MOOC's course i.e. *Developing Your Research Project* as shown in Figure 4. When the course is first announced, 9855 learners joined the course but only 5086 of them visited the course page after the course had started. Noteworthy that less than half of the registered participants had visited the course at least twice to complete some learning steps.

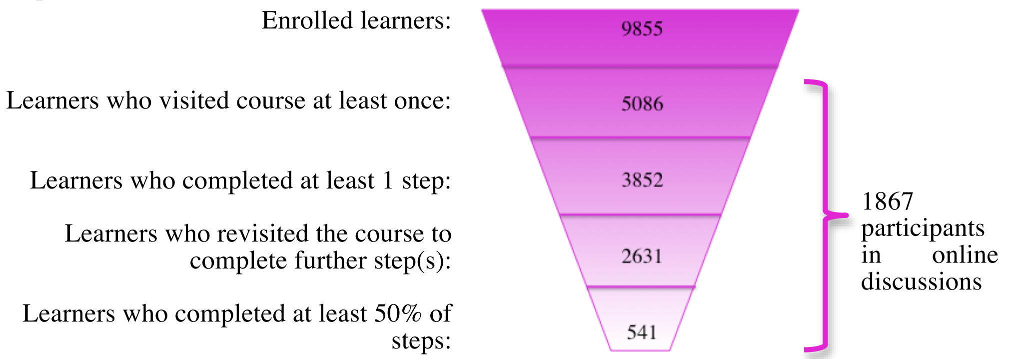


Figure 4. Funnel participation in DYRP MOOC 2014

Although there were 5086 participants who visited at least once, only 1867 actually contributed to online discussions. As it is already explained in Subsection 2.1, learners are able to contribute to online discussions with comments. These comments could be individual comments or replies to a previous comment. Replies identify interactions between learners. The numbers of comments and interactions are both presented in Figure 5. The left circle in shows the number of total comments and interactions among the participants in the DYRP course, while circle on the right presents the number of learners who contributed and interacted with other learners in the discussion sessions. The illustrations denote that while 1867 learners contributed to online discussions by posting at least one comment, only less than half of them replied to the comments.

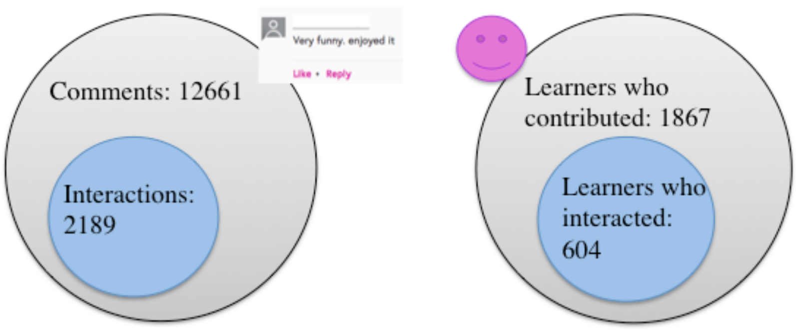


Figure 5. Participations to the online discussions during the course.

In addition to the number of comments and interactions, the number of recurrent interactions in a week is also analysed. Table 2 shows the dramatic decrease on learners' weekly contributions. Correspondingly, the number of interactions and the repliers are at the highest level in the first week, and then it gradually reduces (also see Figure 7). The last column shows repeated interactions in the week. The recurrent interactions identify more than one interactions took place between the same two learners in a week. The number of recurrent interactions also decreases in time.

Table 2: Weekly data on contributions to online discussions.

Data / Week	Number of learners who contributed	Number of comments	Number of learners who replied a comment	Number of replies (interactions)	Number of recurrent interactions
1	1707	5488	447	933	60
2	733	2715	255	708	51
3	386	1247	89	175	11
4	334	748	77	125	4
5	270	689	44	71	5
6	204	516	30	63	8
7	239	596	42	63	3
8	232	662	30	51	1

Figure 6 shows recurrent interactions in each week and also over the weeks. For example, while only 1 learner (see Table 2) repeatedly interacted with somebody else only in the week 8, 3 more learners (see Figure 6) interacted again in the week 8, who are already interacted at least once in the previous weeks. Each number on the axis represents the anonymised learners' IDs. As it is clearly seen, the number of repeated interactions in a week is the highest in the first week. Additionally, it is observed that the number of interactions that occurred over several weeks is extremely low. In other words, the majority of interactions is one-time interaction and does not evolve to a continuous friendship that they can benefit from each other's experiences over the weeks.

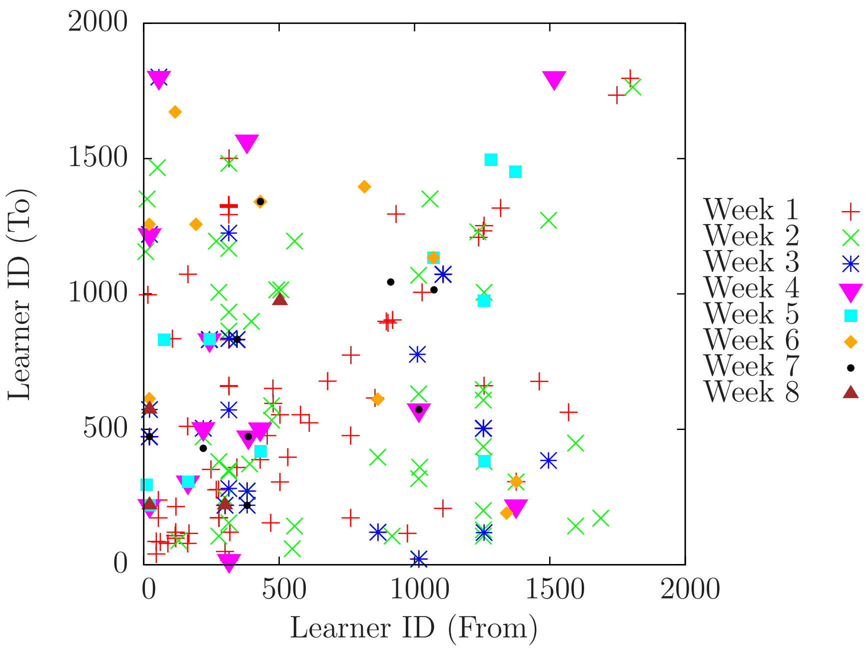


Figure 6. Recurrent interactions over the weeks.

As it is shown in Table 3, low number of learners has repeated interactions (19: 1.75% of all commenting learners) over the duration of the study. According to the provided MOOC dataset by the

FutureLearn, while the average number of comment per learner is 6, it is almost 10 times higher than the general average for the learners with recurrent interactions. Also, while a learner completed 30.5% of the total course steps in average; the learners with repeated interactions completed almost 80 steps of 85 available steps, which is 94.1% of them. In other words, learners who involved in recurrent interactions actively participated all weeks and completed the MOOC.

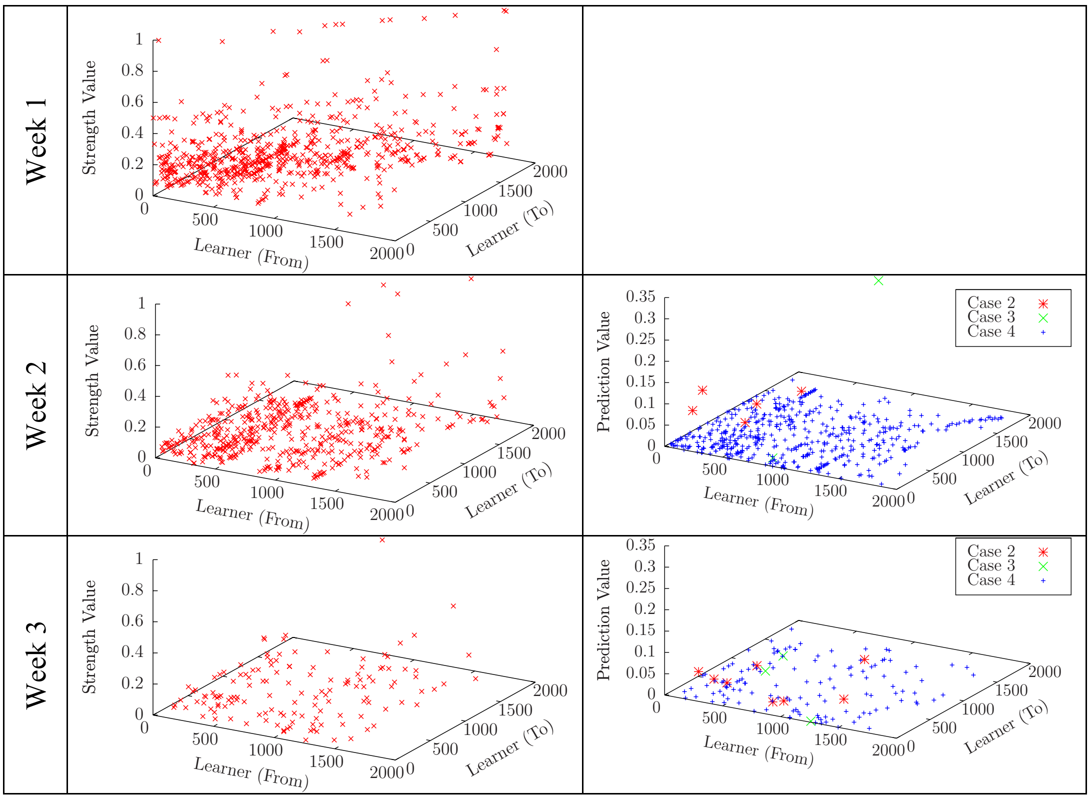
Table 3: Comparison of commenting learners and learners with repeated interactions.

	Commenting learners	Learners with repeated interactions
Total number from original 5086 participant did at least one step	1086	19
Average comments per learner	6	58
Average number of steps per learner	26	80

Moreover, despite the low number of recurrent interactions, their interactions have a pattern. As shown in Figure 6 when an interaction occurred, it is more likely recur in the immediate week. Even though this analysis does not directly prove that learner will finish the course if learners interact repeatedly over weeks, the results are promising. This pattern along with the presented data analysis support the hypothesis that if learners moved forward in the course and continue interacting with each other, there could be more recurrent interactions and learners would engage in the course more and finally complete the course.

4.2. Analysis of the Prediction Model

In order to test the proposed method for predicting interaction, the actual interactions are compared with the prediction results for eight subsequence weeks of the DRYP MOOC's course. The strength and prediction values for each week are illustrated in Figure 7. The graphics on the left shows the strength of friendship between two learners throughout the week. The graphics on the right column illustrates the prediction value for possible occurrence of friendships in the following week once the current week is completed. Learners' situations according to the cases (Case 2-3-4) that are explained in the Section 4 are shown in the graph using colour code.



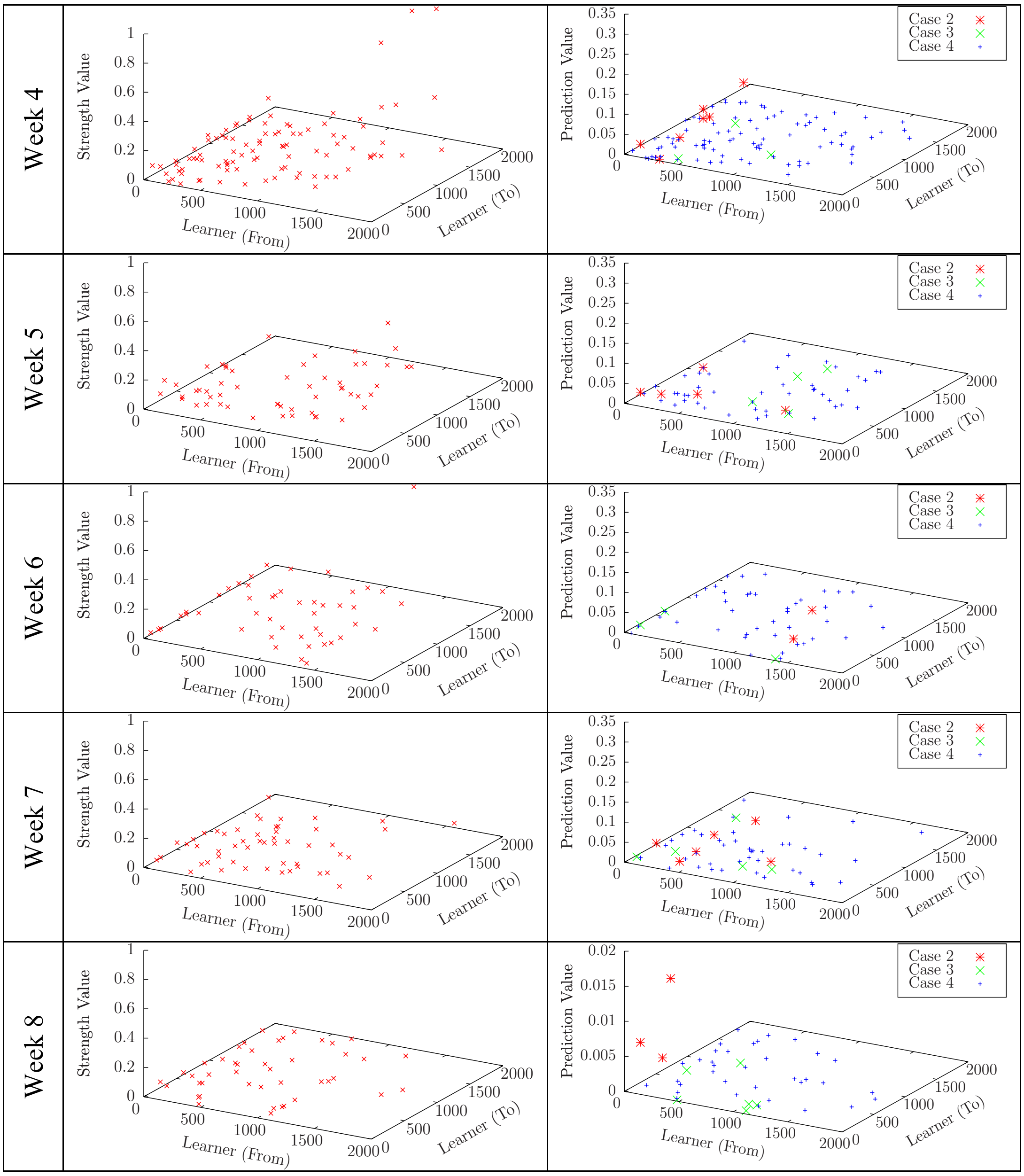


Figure 7. Strength of interactions among learners on weekly basis (left column) and prediction values of possible interactions among learners (right column)

According to the results, our prediction method gives successful results for situations in every case. According to the predictions for *persistent friendships* (Case 2) and *indirect friendships* (Case 3), the probability of occurrence of an interaction is relatively higher if these two learners have had interactions before or have mutual friends. For example, in Week 4, the method predicts possible interactions for learners who have *persisted* and *indirect friendships*. These learners get interacted in real and have relatively higher friendship strength value. However, predicted (possible recurrent) interactions in Week 8 were not happened and those friendships' strength values remained low. This result shows that if those learners are encouraged to interact with each other, they probably could interact with each other and have stronger friendships.

Regarding of the learners that are categorised in *isolated friendships* (Case 4), the probabilities of occurrence of the interactions are pretty low, which are nearly 0. The reason of the low values is that the method assigns high possibility if there is a direct friendship or a friendship through mutual friends. However, since the real strength of friendship values are also low in comparison to others, the results for *isolated friendships* is also consistent. Negatively, even though the method predicts some interactions could happen, some of those interactions are never observed between learners and vice versa. Moreover, the method does not yet able to predict possible interactions of learners who never contributed to discussion board. For example, it is noticeable from the graphs that there are several interactions occurred in Week 2, 3 and 4 that have not been predicted. In order to obtain more precise predictions, the method should be further improved. In order to achieve this aim, further data about learners may be required such as activity history in the course and past activities in other course on the platform if there is any.

5. Conclusion and Future Work

Our data has demonstrated that most of the participations in online discussions are one-time posting. Hence, interactions between learners are remarkably low in comparison to number of comments posted to the online discussion board. We also found out that learners who actively completed most of the learning steps were also those who actively joined online discussions. Their interactions' pattern shows that if learners interacted with each other once, it appears likely that they will interact again in subsequent weeks. Further work needs to be conducted to investigate whether learners who are encouraged to interact with other learners within their social learning network or through mutual friends would have continuous interactions over the weeks and this would eventually improve the retention rate. We are going to propose to build a recommender system for developing learners' social learning networks based on a further developed prediction method for this. For further improvement of the approach, more data about learners such as their other social learning networks on other MOOCs' courses, and their learning history may be required for better prediction of friendships, especially for *isolated friendship*.

References

Clow, D. (2013, April). MOOCs and the funnel of participation. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 185-189). ACM.

Jiang, S., Warschauer, M., Williams, A. E., ODowd, D., & Schenke, K. (2014). Predicting MOOC performance with week 1 behavior. In *Proceedings of the 7th International Conference on Educational Data Mining*.

Khalil, H., & Ebner, M. (2014, June). MOOCs completion rates and possible methods to improve retention-A literature review. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Vol. 2014, No. 1, pp. 1305-1313.

Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2014). Predicting MOOC dropout over weeks using machine learning methods. *EMNLP 2014*, 60.

Margaryan, A., Bianco, M., & Littlejohn, A. (2015). Instructional quality of Massive Open Online Courses (MOOCs). *Computers & Education*, *80*, 77-83.

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International journal of instructional technology and distance learning*, 2(1), 3-10.

Stein, R. M., & Allione, G. (2014). Mass attrition: An analysis of drop out from a Principles of Microeconomics MOOC. *PIER Working Paper*, No. 14-031. Available at http://ssrn.com/abstract=2505028.

Yang, D., Wen, M., & Rose, C. (2014). Peer influence on attrition in massive open online courses. *Proceedings of Educational Data Mining*.

Yang, D., Piergallini, M., Howley, I., & Rose, C. (2014). Forum Thread Recommendation for Massive Open Online

Courses. In Proceedings of Educational Data Mining.

Yang, X., Guo, Y., Liu, Y., & Steck, H. (2014). A survey of collaborative filtering based social recommender systems. Computer Communications, 41, 1-10.

Yuan, L., MacNeill, S. & Kraan, W. (2008). Open educational resources–opportunities and challenges for higher education. Technical report. Joint Information Systems Committee (JISC) CETIS.

Zhuhadar, L., & Butterfield, J. (2014). Analyzing Students Logs in Open Online Courses Using SNA Techniques. In *Proceedings of AMCIS 2014*.