Comparative study of MOOC forums: Does course subject matter?

Anastasios Ntourmas, Nikolaos Avouris, Sophia Daskalaki, Yannis Dimitriadis

Βιβλιογραφική αναφορά:

Ntourmas, A., Avouris, N., Daskalaki, S., & Dimitriadis, Y. (2022). Comparative study of MOOC forums: Does course subject matter?. *Συνέδρια της Ελληνικής Επιστημονικής Ένωσης Τεχνολογιών Πληροφορίας & Επικοινωνιών στην Εκπαίδευση*, *1*, 475–482. https://doi.org/10.12681/cetpe.4338

Comparative study of MOOC forums: Does course subject matter?

Anastasios Ntourmas[1], Nikolaos Avouris[1], Sophia Daskalaki[1], Yannis Dimitriadis[2]

a.ntourmas@upnet.gr, avouris@upatras.gr, sdask@upatras.gr, yannis@tel.uva.es
[1] University of Patras, Patras, Greece
[2] Universidad de Valladolid, Valladolid, Spain

Abstract

Discussion forums are considered important tools for promoting learning in Massive Open Online Courses (MOOCs). They involve learner to learner and learner to instructor discussions. An interesting research question is related to the effect of the subject matter on the characteristics of social networks of MOOCs' discussion forums. Being motivated by Social Centrality Theory, this study aims at further enhancing our understanding of how social centrality is associated with the learners' course achievement within courses in different subjects (e.g. technology vs. humanities). To achieve this, we performed analysis of the discussion forums of two MOOCs of similar size and educational context, one on Programming and the other on History. The study revealed that centrality of learners played more important role in the technology course than the humanities course in relation to course achievement.

Keywords: Massive open online courses; discussion forum; Social network analysis; learning outcomes

Introduction

Massive Open Online Courses (MOOCs) play an increasingly important role in life-long learning and higher education. However, there is a criticism on the pedagogical philosophy most MOOCs are based on. It has been observed that in most cases the salient part is video lectures, which basically perpetuate live lectures, a practice that remains ubiquitous in higher education despite empirical evidence of its low effectiveness (Siemens, Dillenbourg, Fischer, McNamara, Rummel, 2014). In a study on instructional quality for MOOCs (Margaryan, Bianco, & Littlejohn, 2015) it was found that active learning pedagogies based on collaborative activities and contribution of collective knowledge were lacking from most reviewed courses. The most widely used practice for active learning in MOOCs is participation in discussion forums and encouragement for doing so, should be an integral part of learning design. Therefore, an analysis of activities in MOOCs forums can provide insights on the learning effect and thus direct the design towards better pedagogies. A particular question is to study the effect of forum activity on student achievement and whether this effect varies accross different subject matters.

In order to investigate this question, we collected data from the discussion forums of two courses in different subjects, one related to technology (Introduction to Programming) and the other in humanities (World History of Religion) from Mathesis, a prominent MOOCs platform in Greek language, based on OpenEdX technology. We performed social network analysis (Carolan, 2013), in order to identify relationships and interactions among users within the networks that were formed. This approach has been used by Dawson, Gašević, Siemens, & Joksimovic (2014) and Dowell et al. (2015). The objective of these studies was to identify models for the structural properties of the networks and the position of learners in

the network, using four centrality measures, i.e. Degree Centrality, Closeness Centrality and Betweenness Centrality (Freeman, 1978). Other studies investigated the relation of centrality measures to achievement, or course retention (Goodreau, S. M., Kitts, J. A., & Morris, 2009; Joksimović et al., 2016; Margaryan et al., 2015; Siemens, Dillenbourg, Fischer, McNamara, & Rummel, 2014).

Methodological framework

The aim of this current research was to perform analysis on the networks that are constructed from student activities in the forums, and answer the following research question: In which way learners' engagement, as described by their centrality values in the forum networks, relate to their course outcome, and how this is related to the two different subject matters?

In order to address the research question, firstly we built social networks derived from the discussion forums. Each network is a directed weighted graph representing the interactions within a single thread. A basic assumption for the construction of the graphs was that of Direct Reply (Wise, Cui, & Jin, 2017), i.e., each message posted in a thread is considered to be a reply directed to the previous one. For example, if author B posts a message directly after author A, then a directed edge B→A will be added to the graph.

For these networks, global network metrics (density, clustering coefficient and reciprocity) can be derived. *Network density* is an indicator of group cohesion and measures the portion of existing connections compared to the 'potential' connections that could exist between any two users in the thread. The *clustering* coefficient is an indicator of the network formations and the users' connection patterns. It measures the degree to which users in a thread tend to cluster together by keeping replying to previous posts, and continue the discussion on the subject of the thread. Similarly, *reciprocity* measures the users' tendency to cluster together by forming mutual links.

In addition to the global network measures, we also need to calculate node centrality measures for all students that participated in the forums. Such measures are *Degree*, *Closeness* and *Betweenness Centrality*. *Degree Centrality* is the total number of direct connections a user has within the network. Since the networks are directed, we need to distinguish *in-degree* (replies received) and *out-degree* (replies send) *centrality*. This distinction is undertaken in order to identify the more 'active' users who had the tendency to answer to many posts and the 'passive' ones who had the tendency to just post and wait for replies. *Closeness Centrality* is calculated as the reciprocal of the sum of the lengths of the shortest paths between the selected node and all the others in a network. This definition means that the closer a node is to the other nodes, the more central appears to be. Lastly, *Betweenness Centrality* represents the degree of which a node stands between shortest connections of others.

In order to find the association between Centrality measures and learning achievement, measured by the successful completion of the course, we need to perform hypothesis testing between the group of users who earned the course completion certificate and those who did not, for each course separately. In the next section, application of the described methodology is outlined and discussed.

Case study: Analysis of two MOOCs

Description of data

The data used for this study were retrieved from two MOOCs offered in 2017 on the mathesis.cup.gr platform. The first course, '*Introduction to Python*' (C1), aimed to introduce learners to computer programming through Python. The second course, '*World History: Man*

*versus the Di*vine' (C2), aimed to introduce learners to the history and evolution in time, of Asian religions during the Second Circle of World history. The duration of the two courses was 6 and 9 weeks, respectively. Their learning design consisted of video lectures, assignments and weekly tests. Students were also encouraged to use the course discussion forum, organized in discussion threads. Participation in the forums was optional. Learners could create their own threads in the forum or participate in existing threads, created either by the instructors or other learners, by posting replies. The data used for the analysis contained users' demographics, grades from their assignments, enrollment status and their contributions in the forums. Data were also collected from two questionnaires in each course, conducted, one at the start (users' motivation and background) and the other at the end (users' experience and evaluation of the course).

Analysis of data

According to enrollment data, 7614 students enrolled for course C1 and 5569 for C2. From those 28.48% in C1 and 54.24% in C2 participated in the activities of the two courses, i.e. posted at least one message in the forum or submitted a test or assignment.
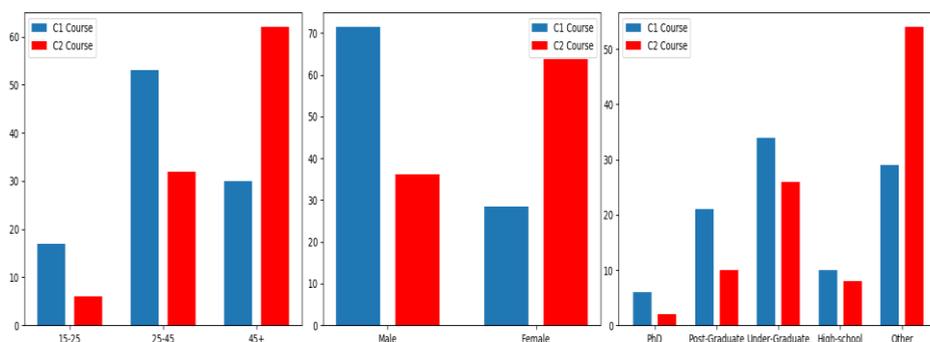

Figure 8: Distribution of engaged users according to age, gender & educational level

Figure 1 depicts the demographics of the students that participated in the two courses (y-axis: percentage of users). The students of the technology course (C1) were younger, mostly male, and with higher educational level compared to the humanities ones (C2).

A common grading scheme was used for both courses (0 to 1 scale). Successful course completion required a final grade of at least 0.5, calculated as the average of grades in assignments and weekly tests. In C1, out of all 2169 participating users almost half (1154 users) successfully completed the course (53.2%). In C2, out of 3021 participating users, 2500 successfully completed the course (82.7%).

A detailed view of the final grade distribution for the participating students of C1 and C2 is shown in Fig. 2(a) and 2(b). In C1 a significant percentage of students (34.2%) achieved the maximum grade (1), while a smaller percentage (14.8%) achieved grades between 0.6-0.8. . On the other hand, in C2, the largest portion of users (68.5%) achieved grades over 0.8 and a much smaller proportion (19%) between 0.6-0.8. This difference can be explained by the fact that assignments for C1 require programming, so normally, if the submitted code was correct, the maximum grade was achieved. Assignments for C2, on the other hand, require expressing opinions and judgments on historical matters so grading is more subjective.

The discussion forums for the two courses are structured in three levels, threads, posts, and replies. The total number of threads and posts are presented in Table 1. The two courses are of similar size, in terms of forum activity.

Table 1: Engaged students, threads and posts in forums

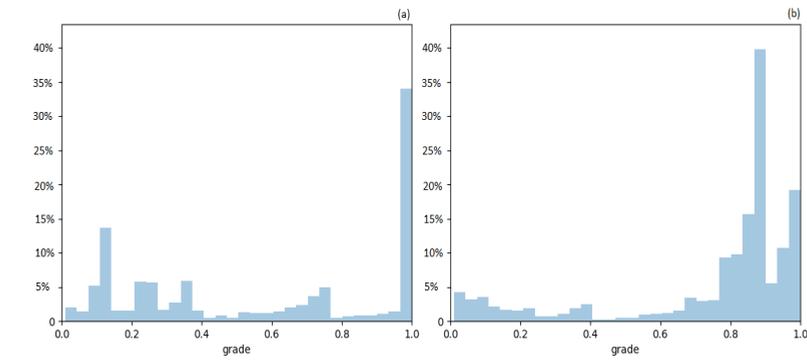|	C1 Course	C2 Course
# Students Engaged in forum	641	564
Total # Threads	380	387
Total # Posts	3281	3765

Figure 2: Final grades for users of (a) course C1 and (b) course C2
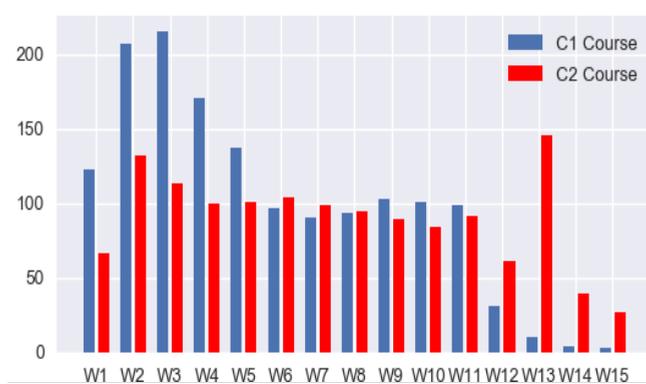

Figure 3: Students participating in the forum per week

Fig. 3 shows forum activity distribution per course week. Lectures for courses C1 and C2 were held for 6 and 9 weeks, respectively, forum activity however was extended after the end of lectures, relating to capstone project preparation and discussion, extending up to week 15.

Patterns of participation differ in C1 and C2. Students in C1 were very active during the first five weeks, their number exceeding 150. This high participation was due to technical

problems they faced. Later the number of participating users leveled up to approximately 100, while after week 11 it decreased exponentially. For course C2 there was more uniform participation all along, with a significant increase only during the week of the final assignment submission.

Forum analysis

Following analysis of forum networks, 380 graphs were built for C1 and 387 for C2. In Figure 4, thread size distribution is shown for the two courses. It is notable that both courses' discussion forums contain many threads with zero, or just one or two posts. Such threads were discarded, as they do not represent significant group interactions. The significant networks were thus 137 out of the 380 for C1, and 147 out of the 387 for course C2. These networks were considered for the rest of the study.
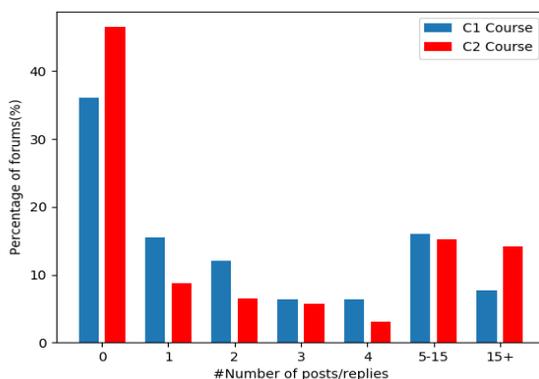

Figure 4: Frequency distribution of number of posts/replies in threads

Table 2: Global network metrics of the social networks extracted from courses' discussion forums

	C1 Course	C2 Course
Mean Network density	0.008	0.012
Mean Reciprocity	0.37	0.40
Mean Node degree	0.23	0.13
Mean Edge weighted degree	0.16	0.22
Mean Clustering Coefficient	0.12	0.21

In Table 2 the global network metrics for significant networks are shown. They indicate great similarity of the networks analyzed, in terms of density and reciprocity. Given the similar number of engaged users, on C1 the average node degree was almost doubled the one of C2, but the average edge weighted degree was higher in C2. These differences suggest users in C1 course interacted with more users, but in C2, users had more repeated interactions with the same users in contrast to C1.

Another notable difference refers to the mean clustering coefficient, which in C2 networks is approximately double to that of C1. This difference can be explained as C1 course is a technology course and many threads concern users seeking technical support. To verify this hypothesis, we inspected several threads' discussions for C1 course. Most of them contained questions on technical matters of the course; so when a question (post within a thread) was answered correctly, no one else carried on the discussion. Except only on specific occasions, where alternative solutions were compared for their efficiency between users. On the contrary, in C2 discussion forum, many threads related to philosophical debates, producing longer dialogue threads and greater familiarity among users. As a result several users when they got replies from other specific users, they tended to communicate again in future posts in a more personal way, in contrast with C1, where communication was mostly formal.

Next we tested whether there were significant differences on centrality measures based on achievement, for the two courses. For C1 for all centrality measures we discovered significant differences between the population that earned the certificate compared to those who did not ($p \ll 0.001$). For C2, we also found significant differences when $a=0.05$. For *In-degree* ($p=0.05$), *Out-degree* ($p=0.04$), *Closeness* ($p=0.008$) and *Betweenness* ($p=0.049$), so the null hypothesis is rejected. Overall, we can claim that there were significant differences between centralities for those who earned the certificate and those who didn't for both courses. However, this difference was more prevalent for course C1. So, for our research question, we may conclude that learners of higher engagement with the forum, as measured through centrality indices, are most probable to successfully complete the MOOC.

Questionnaires reveal that 80% of the users in both courses were using the discussion forum for less than an hour a week and 15% of them between 1-3 hours. There seems to be no difference in the degree of participation in the forum between the two populations. The data from the questionnaires also reveal that the users had different purpose of participation in each course. In C1 the majority of the participants were university students, teachers and scientists, while in C2 most participants were citizens interested in history. C1 is a skill-oriented course and C2 aims at acquisition of historical knowledge. This may explain the different motivation for participation in the course and furthermore in the discussion forum. In C1's forum, users search for solutions to the problems they face in programming, while in C2 they seek opportunities to discuss course issues with other users.

Demographics also show important differences between the student populations of the two courses (Fig 1). They reveal that the education level of C1's users was mostly under-graduate, post-graduate and doctoral, while in C2 54% did not state their education level. C1 course consisted of a younger audience (25-45) mostly male (71.47%), while C2 consisted of an older audience (45+) with more female users (63.73%). The community formed in each course's discussion forum is of different type and this may reflect to the different impact centralities may play on course achievement.
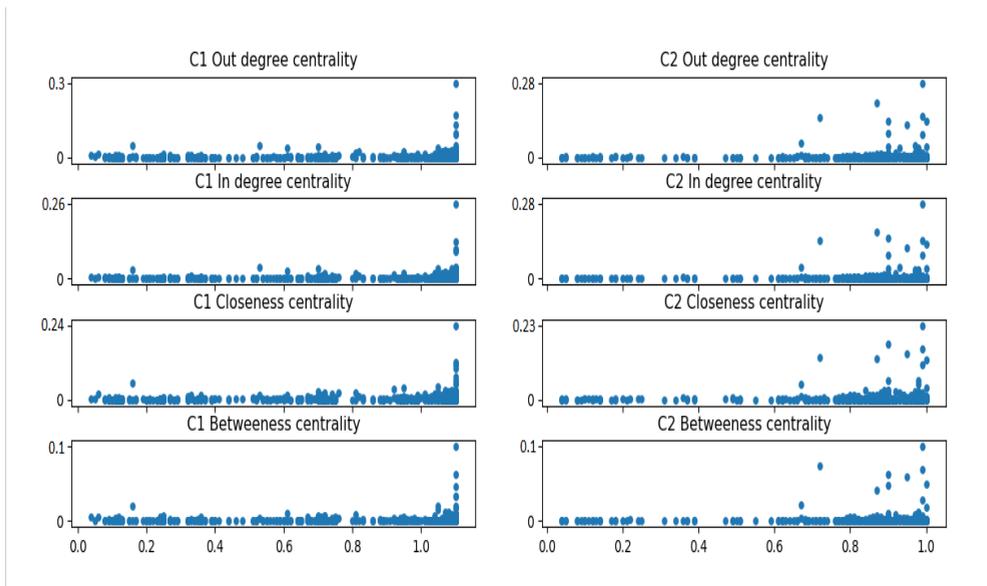
Figure 5: Scatter plots for centrality vs final grade for the two courses

As a last step in Fig. 5 we plotted all four centrality measures (y-axis) against final grades (x-axis) for the users of both courses. It can be observed that "typical" values of centralities in the forum networks is not related to better grade in the course. However, centralities that are much higher than average are related with higher grades.

Conclusions

In this study, we performed social network analysis on the discussion forums of two different MOOCs. We used network centrality measures to identify the position of a student in a given forum. The first aim was to investigate if the centrality measures of a learner relate to achievement. In addition, we investigated if the centrality of a learner is related with the final grade. It was found that in both the technology and humanities courses, the centralities of learners who earned the certificate was significantly higher than those who did not. The difference was more significant for the technology course. This leads to the the observation that in the specific MOOC communities, learner network centrality is related to achievement.

Further content analysis of the discussion forums and the demographics of the engaged students revealed important differences, which may provide some explanations for the different behavior of forum participants in the two courses. The forum participation distributions (Fig. 5) revealed a big number of empty threads for each course. These threads were created mostly by users and not by moderators/instructors. They included often technical questions that could had been answered in other existing threads. Forum threads had different characteristics in the two courses, due to the different nature and use of the forum in the two courses. An interesting aspect to further investigate, is the relation of instructional design in the use of the forum in the different courses and subject matters. A direction that is proposed by learning science experts (Fischer, 2014) is that of a social constructivist approach that involves more engagement of students in the forum in discussing

open questions and collaboratively solving problems. These approaches need to be further investigated in the context of MOOCs.

In future research, we need to perform the conducted analysis to other courses with similar subject matters in order to confirm the findings reported in this paper, while more extensive analysis of content and forum participant roles can further reveal different patterns of interaction in different subject matters.

Acknowledgements

References

Carolan, B. V. (2013). *Social Network Analysis and Education: Theory, Methods & Applications*. SAGE Pub.

Dawson, S., Gašević, D., Siemens, G., & Joksimovic, S. (2014). Current State and Future Trends: A Citation Network Analysis of the Learning Analytics Field. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (pp. 231–240). New York, NY, USA: ACM. https://doi.org/10.1145/2567574.2567585

Dowell, N. M., Skrypnyk, O., Joksimovic, S., Graesser, A. C., Dawson, S., GaĹˇevic, D., … Kovanovic, V. (2015). Modeling Learners' Social Centrality and Performance through Language and Discourse. *International Educational Data Mining Society*.

Fischer, G. (2014). Beyond hype and underestimation: identifying research challenges for the future of MOOCs. *Distance Education*, *35*(2), 149–158. https://doi.org/10.1080/01587919.2014.920752

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239.

Goodreau, S. M., Kitts, J. A., & Morris, M. (2009). Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography*, *46*(1), 103-125.

Joksimović, S., Manataki, A., Gašević, D., Dawson, S., Kovanović, V., & de Kereki, I. F. (2016). Translating network position into performance: importance of centrality in different network configurations (pp. 314–323). ACM Press. https://doi.org/10.1145/2883851.2883928

Margaryan, A., Bianco, M., & Littlejohn, A. (2015). *Instructional quality of Massive Open Online Courses (MOOCs)* (Vol. 80). https://doi.org/10.1016/j.compedu.2014.08.005

Siemens, G., Dillenbourg, P., Fischer, G., McNamara, D., & Rummel, N. (2014). Where are the Learning Sciences in the MOOC Debate? (pp. 15–17). Presented at the Learning and Becoming in Practice: 11th International Conference of the Learning Sciences (ICLS), Boulder.

Wise, A. F., Cui, Y., & Jin, W. Q. (2017). Honing in on Social Learning Networks in MOOC Forums: Examining Critical Network Definition Decisions. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 383–392). New York, NY, USA: ACM. https://doi.org/10.1145/3027385.3027446