

Data values on the MOOCs in the university's educational : 80

¹Z. Harmouch, ²K. Ghoulam, ³B. Bouikhalene and ⁴H. Mouncif

^{1,2} Department of Mathematics and Informatics, Sultan Moulay Slimane university
FST Beni Mellal Morocco

^{3,4} Department of Mathematics and Informatics, Sultan Moulay Slimane university
FP Beni Mellal Morocco

Abstract : *The term MOOC (Massive Online Open Course) is used to refer to online teaching platforms, offering open courses which can apply to hundreds, thousands or even tens thousands of students simultaneously. So, more than ever, the amount of data onto learners, teachers and courses has exploded to the world of MOOC. In addition, more data are available for learners from several sources, including social networking platforms like Facebook, Twitter, Youtube etc. The purpose of this paper is not analysing the MOOC phenomenon in general, but to give an original vision on the MOOC, through its students, their profiles, and their activities during the course. In order to cope with this massive data, we will use a new technology "Big Data Analytics" to explore and analyse the behaviour of learners, their profile, and their activities during the course. For this, we analyse the data extracted during a course, launched on the educational platform of the University Sultan Moulay Slimane. We describe the communities of students, their socio-economic profiles, their motivations, their activities on the Facebook group created for this course, we will also study how the exchanges on this social network structure during the course.*

Keywords : MOOC, Big Data, social networks, Facebook, online education.

1. Introduction

The craze massive platforms that online courses have generated, made MOOC an object of attention : public institutions, national and international, as well, that many schools, universities, and large companies are investing in the creation of this type of course, with various objectives : help in the development of continuing education, the formation of the developing countries, increase their visibility etc. the growing literature on the MOOC, attests to the great relevance of the subject, that these publications are targeted at the general public or a more specialized public (see, for example, Rolfe, 2015 for a recent review of the literature [1])

The aim of this paper is not to provide an analysis of the phenomenon MOOC, its challenges, its successes, its failures and its future, but to offer an original vision of a MOOC, through its students, their profiles, and their activities during the course.

Indeed, social networks are often described as an essential ingredient of an effective online course [20], because this asynchronous communication is the main educational interaction on this type of

course. Many studies ask the question of the use of social networks, the results of these various studies are sometimes contradictory, So, the goal of this article is to make a contribution to this debate, in order to provide elements for a better use of social networks or to think about alternative solutions. Our job is to analyze - from the study of the conversation threads, comments, and messages posted or sent on the Facebook group of a MOOC - the behaviour of various participants, in order to provide a diagnosis, allowing forward a more efficient use of these communication tools, between students, and between students and teaching team, with the Big Data Analytics technology. Note that this analysis is, to our knowledge, one of the first analysis of student behavior on a Moroccan MOOC : This study therefore also value comparing with previous studies, mainly performed on the participants of large international platforms [2] (Brinton and al., 2013) and participants of our platform.

The rest of this paper is organised as follows : in Section 2, we will briefly present the course "Information Processing", to locate the context of our study, and give a descriptive overview of the

profile of students who have registered in this MOOC. In Section 3, we will focus on the statistical analysis of data collected on the Facebook group by focusing in particular on the community trade patterns.

Finally, we point out in the introduction that the anonymity of students has been strictly respected: the data presented here are for student profiles and activities, without the possibility of identifying them.

2. Overview of the MOOC "Information Processing"

The educational platform of Sultan Moulay Slimane University is the official platform to promote the use of online courses in the University. The course "Information Processing" was part of the initial formation UTD-HRM (University Technology Diploma - Human Resource Management) It started February 8, 2016, for a period of 5 weeks. This course is intended to provide an introduction to basic concepts in information processing.

The content of the course was structured in five main sessions, corresponding to the 5 weeks of the course, and unveiled gradually. The concepts discussed were:

- Week 1: IS (Information System) and HRMS (Human resource management system)
- Week 2: Conception of an IS and Introduction to Merise method;
- Week 3: CDM (Conceptual DataModel);
- Week 4: LDM (Logical Data Model)
- Week 5: MPD (Physical Data Model)

Each week itself was structured into short sub-sequences each having a video of twenty minutes and a self-assessment quiz on the concepts already seen in the video. In addition, each set of sequences was accompanied by an introduction which specified goals of the week, the prerequisites and the time required to, firstly, watch videos, and secondly, realize personal work. A synthesis part concludes each week, retaking deepening the concepts that we discussed on certain aspects for those wishing to go further. Finally, a duty is requested at the end of the third, fourth and fifth weeks to practice the realization of deferent data models.

In total, the course requires an average weekly workload estimated at approximately 5 to 7 hours per week, that is about 30 to 42 hours of availability on 5 weeks. Finally, in parallel with the course tools set to the disposal of students, the course contained features to allow the exchange between students and between students and teacher : next to wiki and forum, a group on the Facebook social network was created and provides to students of the objective MOOC.

From the viewpoint of pedagogical choices, the aim of the course was to present to an audience totally novice, the information processing concepts required to conduct an elementary data analysis. The course was oriented to present the aspects of data design and modeling based on Merise method, and illustrate them in cases of simplified and generic studies.

MOOC 'Information processing' got 1622 registrants whose 932 were already registered on the first day of the course. At the end, 70 students have completed all of the five exercises. An auto declaration questionnaire offered to students during their registration, allowed us to obtain information on their profiles: 85% of registrants (1403 students, of which 32% women) finished the course, making it a source of data that can be considered reliable. Among the highlights, can be noted that about three-quarters of students, whose it was possible to find their geographical location, were domiciled in Morocco, the others mostly foreign students from francophone Africa. The MOOCs, open and free, are just required to have access to the internet, indeed they have a particular interest in learning in the third world countries [3] (see eg Cisel, 2014).

Registered students have an average of about 32 years, indicating a relatively small proportion of students are students who wish to deepen or complete a course, they tried to follow it themselves in the classical way (i.e. initial education at the University e.g. Alario-Hoyos et al. 2013 [4]). This is corroborated by a very strong presence of registrants already having a university degree (almost 53% of responses), that is higher than the level of classical education that the course "Information processing" target.

3. Analysis of Facebook group

One of the tools available to students on this MOOC is the group on the social network

Facebook: it's a place for exchanges between students next to forums and wikis, but also between students and the teaching team, so, it allows to have an image of interactivity during the MOOC. With this tool, any person, whether a student or teacher, can begin a discussion thread which corresponds to a new post or message on the Facebook group, any other person may respond to them in the same thread by a comment or message, the number of comments or responses for a given thread is not bounded. The discussion thread has, therefore, as a general rule, a thematic consistency, and as every thread are accessible to all, a particular theme is normally addressed in a single thread, but it is common that a topic is dealt in two or more different thread. When the person who opens a new thread did not bother to check, by reading all of the other threads, if the answer to his question was already present in another one.

In this section, we propose to analyze these discussion threads, by first describing the Group on the social network Facebook from a quantitative viewpoint, and then concerning the structuring of student groups, and finally, analyzing the comments themselves.

4. Descriptive analysis of the Group

The social network group was active from 8 February 2016 to 13 March 2016, it has been used by 122 identified students, two members of the teaching team, and 14 anonymous people, who have opened 151 threads corresponding to 613 separate comments. Therefore, the number of participants in the Group represents a relatively low percentage of the number of registrants (about 7.52%). However, if you compare the number of students having proposed a solution for the five exercises to make (251 students), we find that students active, seeking to invest in the course by participating via Facebook, are twice as many students who have

fully completed the course. These results are consistent and even rather the sign of a quite important activity in the MOOC « Information Processing » if compared to the studies of Breslow et al. [5] and Manning and Sanders (2013) [6] which show that in general less than 10% of registrants (and most often less than 5%) participate in discussion threads.

Previous studies have shown that the role of the instructor or moderator is important for the success of the discussion threads (Anderson and Kanuka [7], 1997;) Berge, 1995 [8]). Similarly, Salmon (2004) [9] based online learning and the interaction of the discussion forums on the fact that the instructor or the moderator encourages discussions and activities (Harasim et al. 1995 [10];) Mak et al. 2010 [11]). Note those other authors are less enthusiastic, showing that few teachers are involved in discussions or even that their role can be negative in terms of student engagement (see Baxter and Haycock, 2014 [12]]; Onah and al., 2014 [13]). On the MOOC 'Information processing', the role of the teaching team was to not open discussion threads to offer the freedom to students to address the themes of their choice: the analysis that follows is, therefore, that of a discussion thread which has not been structured a priori. However, the teaching team has been very active with 217 comments (about 35.38% of total comments) and focused its interventions on a little longer discussion threads (threads very short, 2 or 3 comments, often correspond to a question from a student that another student responds appropriately). In addition, Figure 1 shows the distribution of the number of comments per student. If some students involved frequently on the Facebook group (with a maximum of 16 comments to a student), most students are involved very little: 32 students (more than 26.23% of the group members) have posted a comment and 20 students have posted two.

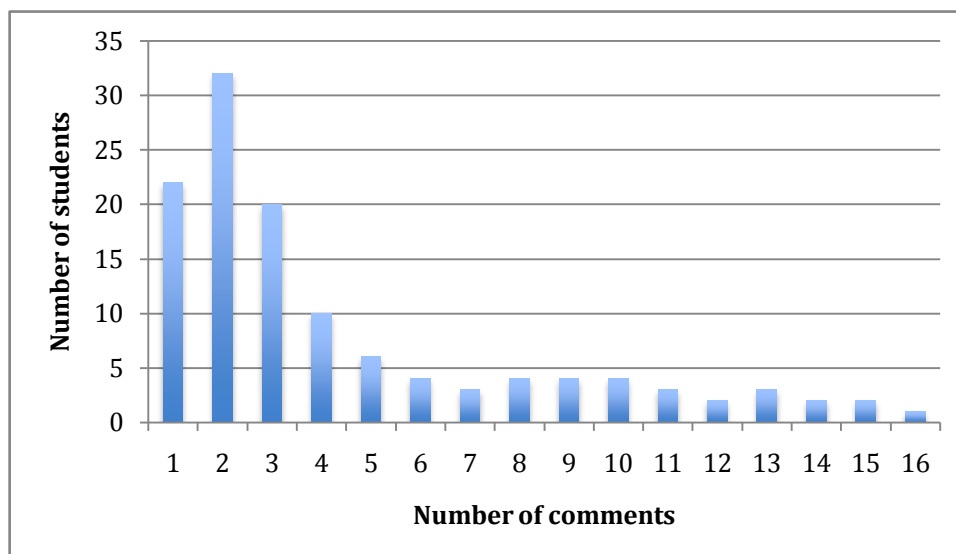


figure 1. distribution of the number of comments posted per student (excluding teaching team).

We see therefore the community be structured around some very active students who are engaged on the Facebook social network and with, a lesser role compared to the teaching team, boosting exchanges. Note also that the profile of students participating in the group differs slightly from the entire students profile: participants in the group are older (the average age and the median age is in the order of 41 years against 27 years for all students) and students with a bachelor's degree or more are more present on the social network group (in particular, 19% of participants in the social network having a bachelor's degree while they represent only about 14% of registrants). These results are in part similar to those found by Gillani and Eynon (2014) [14] which shows that participants in the forums are more graduates. However, in the same study, they appear younger

than all the students, which is a notable difference from MOOC 'Information processing '. One hypothesis to explain this phenomenon could be that the teaching team having left the initiative of discussions, students more experienced took a bigger role, playing the role of moderator within social networks.

The relatively high number of active students in the group compared to students who have completed the course is partially explained by the figure 2, which shows the evolution of the number of comments made during the time. There are clearly identified several phenomena : the first visible is a global tendency to the decrease in activity on the social network, already highlighted in Brinton et al. [13] and easily attributable to abandonments and dropout of students, as the course becomes more difficult.

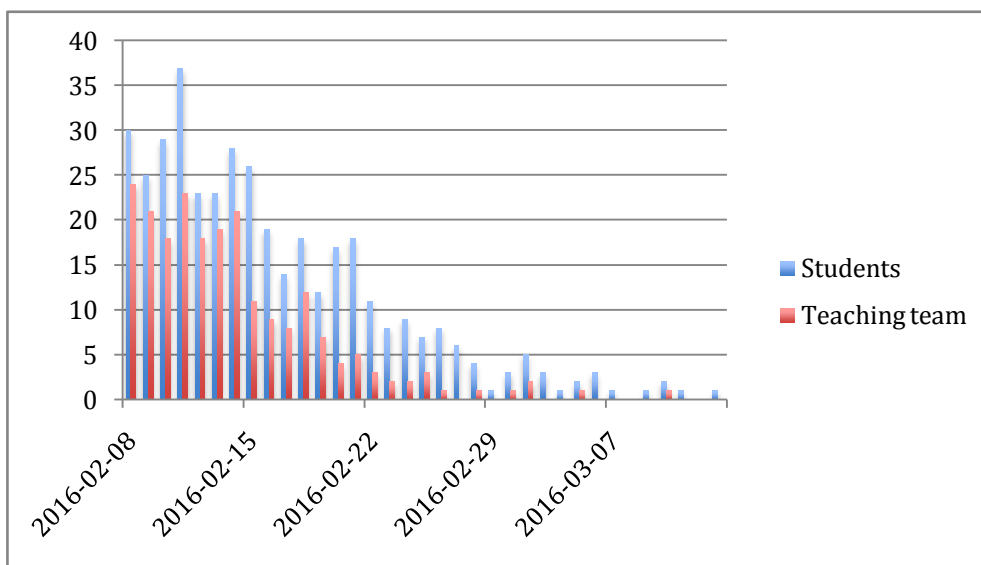


figure2. Evolution of the number of comments during the MOOC

is clearly

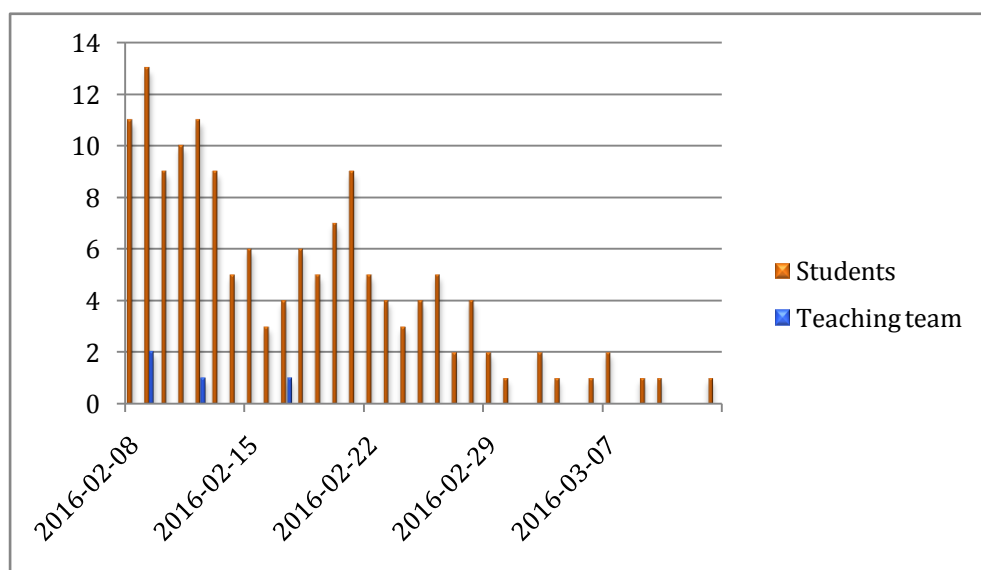


figure 3. evolution of number of threads of discusstssion during the MOOC

Also, it

visible that the appearance of new videos on the platform, is always followed by a rebound of activity on the Facebook group : the five sets of videos are identifiable by an increase in the number of threads discussion and comments on the group. Therefore, we note that, although a significant part of the comments of the first week involves purely technical aspects of using the platform, the number of comments in the fifth week of the MOOC (2nd week of March) is four times less than the first week. In particular, a net dropout operates after the second series of videos (which address the conception of an IS and introduction to Merise method), and to a lesser extent, after the fourth.

Finally, changes in the number of teaching team's comments follow fairly closely that of students. These remarks show that the analysis of the activity on the Facebook group, performed in real time, is, in general, a good indicator of the activity of students, even if a small proportion of those who participates. In particular, they allow the teacher to easily identify which parts of the course least understood, and would, therefore, be used to propose additions or clarifications to students.

5. Communities

The analysis of the group refers to the social network theory that designs social relationships in terms of nodes and links (notions from the Graph Theory). The nodes are usually social actors in the

network, but they can also represent institutions, and links are the relationships between these nodes.

Indeed, it is generally accepted in a natural way (Freeman, 2004 [15]), that human groups are organized into cohesive social subgroups. From the perspective of the mathematical object "graph", the definition of these communities is not completely uniform, and may vary depending on the application domain. However, quite consensually, the notion of communities refers to a partition of the vertices of the graph, such as the groups of vertices are dense (i.e. with a large number of edges within the group) and connected by a small number (comparatively) of edges. We looked for communities in the graph students by maximizing the modularity. It is one of the most classical approaches (Newman and Girvan, 2004 [16]) and we used the method of approximate optimization described in Rossi and Villa-Vaughan (2011) [17]. We have, moreover, tested the significance of this partition by comparing it to a maximum modularity of a set of random graphs, with the same distribution of degrees that as our students graph. The optimal partition that we obtained contained 11 classes (number of vertices) whose numbers are given in table 1 and who had a modularity equals to 0.507, far superior to the maximum modularity found for 100 random graphs (0.364) : The communities of the graph structure is quite marked, unlike what concludes McGuire (2013) [18].

Community	1	2	3	4	5	6	7	8	9	10	11
Number of vertices	10	12	9	5	8	13	4	12	11	10	8

table 1. The number of vertices in each class (community) found by modularity optimization.

So there seems to be, as also point Gillani and Eynon (2014) [14], an organization of students in groups of people who make a habit to discuss more frequently between them. Gilani and Eynon (2014) [14] show that there is a geographic structure within these groups. In the same way, we studied the classes to find differences between profiles of students from the perspective of age, gender, geographic location and education level. Only age was significantly different depending on the community, confirming the results of the tests on the correlation structure of the graph and description of individuals described above. The age distribution by a community is shown in Figure 4. It shows in

particular that the smallest community (7th containing 4 students) has the lowest age (the median age is approximately equal to 30 years) while the community 3 (9 students) has a higher median age, close to 50 years, despite three outliers whose age is less than 20 years. It seems again that the strategy for the teaching team, allow a better communication between the various classes of ages (ensuring, for example, to limit the effects of differential expression or cultural reference) could be a key to power the life of the social network. Another solution might be, instead, to identify and understand the specific needs of these communities to produce personalized speech toward each.

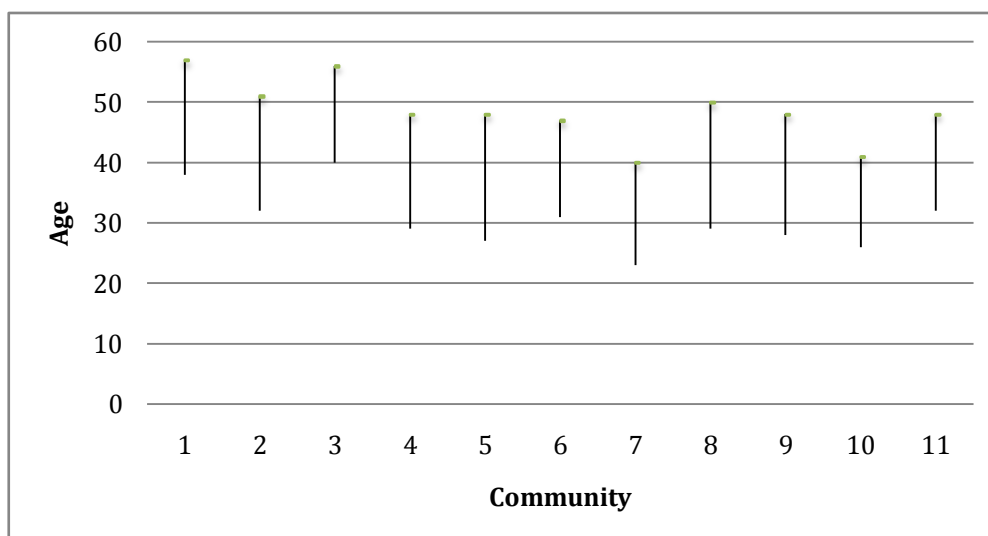


figure3. Distribution of age according to the community

6. Conclusion

This paper presented a quantitative study of the participants of the MOOC "Information Processing" and especially the participants of social network Facebook. The analysis showed that students enrolled in the course have a profile and behavior, that are similar to those identified in previous studies, on the course of the major international platforms. In particular, their profile is quite different from the University student to which this course could address : they are older, already graduated, and a significant number of them followed the course from francophone foreign countries. Activity around the course was mainly produced by a 'hard core' of students who themselves had a very contrast activity (some extremely active and other active). The most active students have proved to be also the older. Unlike other works has shown, and appear to play a role similar to that of the teaching team by providing a

significant portion of responses rather than opening discussion threads.

On the other hand, we have shown that contrary to previous studies, participants are structured into groups of independent exchanges of most features of their profiles, except age. This is a point that should be considered by the teaching team to manage the Facebook group and promote the engagement of students, either by playing a role of a binder or by producing a custom response toward each group.

The activity analysis also reveals a phenomenon of progressive dropout already observed otherwise. This is a good vector to assess which parts of the course were deemed too difficult. It would be interesting to correlate these activity measures to exercise participation and success to duty, but the available data do not allow for a clear link between students on the platform and the proposed external resources, as pointed out by Anderson et al. [19] (2014) in their study.

References

1. Rolfe, V. (2015). A systematic review of the socio-ethical aspects of massive online open courses. *European Journal of Open, Distance and E-Learning*, 18(1), 53-72.
2. Brinton, C. G., Chiang, M., Jain, S., Lam, H., Liu, Z. et Wong, F. M. F. (2013). Learning about social learning in MOOCs: from statistical analysis to generative model. *IEEE Transactions on Learning Technologies*, 7(4), 346-359. doi:10.1109/TLT.2014.2337900
3. Cisel, M. (2014). Analyzing completion rates of first French xMOOC. Dans U. Cress et C. Delgado Kloos (dir.), *Proceedings of the European MOOC Stakeholder Summit 2014* (p. 26-32). Barcelone, Espagne : P. A. U. Education.
4. Alario-Hoyos, C., Pérez-Sanagustín, M., DelgadoKloos, C., Parada-Gélvez, H. A., Muñoz-Organero, M. et Rodríguez-de-las-Heras, A. (2013). Analysing the impact of built-in and external social tools in a MOOC on educational technologies. Dans

- D. Hernández-Leo, T. Ley, R. Klamka et A. Harrer (dir.), *Scaling up Learning for Sustained Impact : Proceedings of the 8th European Conference on Technology Enhanced Learning* (p. 5-18). Berlin, Allemagne : Springer
5. Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D. et Seaton, D. T. (2013). Studying learning in the worldwide classroom: research into edX's first MOOC. *Research & Practice in Assessment*, 8(1), 13-25.
6. Manning, J. et Sanders, M. (2013). How widely used are MOOC forums? A first look [billet de blogue]. Retrieve on April 22, 2015 from blog Signals – Thoughts on Online Learning : <http://signalblog.stanford.edu>
7. Anderson, T. et Kanuka, H. (1997). On-line forums: new platforms for professional development and group collaboration. *Journal of Computer-Mediated Communication*, 3(3), 0. doi:10.1111/j.1083-6101.1997.tb00078.x
8. Berge, Z. L. (1995). Facilitating computer conferencing: recommendations from the field. *Educational Technology*, 35(1), 22-30.
15. Freeman, L. (2004). *The development of social network analysis: A study in the sociology of science*. Vancouver, Canada : Booksurge.
16. Newman, M. E. J. et Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2). doi:10.1103/PhysRevE.69.026113
17. Rossi, F. et Villa-Vialaneix, N. (2011). Représentation d'un grand réseau à partir d'une classification hiérarchique de ses sommets. *Journal de la société française de statistique*, 152(3), 34-65.
9. Salmon, G. (2004). *e-moderating. The key to teaching and learning online* (2e éd.). Londres, R.-U. : RoutledgeFalmer.
10. Harasim, L., Hiltz, S.R., Teles, L. et Turoff, L. (1995). *Learning networks*. Cambridge, MA: The MIT Press.
11. Mak, S. F. J., Williams, R. et Mackness, J. (2010). Blogs and forums as communication and learning tools in a MOOC. Dans L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. De Laat, D. McConnell et T. Ryberg (dir.), *Proceedings of Networked Learning Conference* (p. 275-284).
12. Baxter, J. A. et Haycock, J. (2014). Roles and student identities in online large course forums: implications for practice. *The International Review of Research in Open and Distance Learning*, 15(1), 20-40.
13. Onah, D. F. O., Sinclair, J. E. et Boyatt, R. (2014). Exploring the use of MOOC discussion forums. Dans *Proceedings of London International Conference on Education* (p. 1-4).
14. Gillani, N. et Eynon, R. (2014). Communication patterns in massively open online courses. *The Internet and Higher Education*, 23, 18-26. doi:10.1016/j.iheduc.2014.05.004
18. McGuire, R. (2013). Building a sense of community in MOOCs. *Campus Technology*, 26(12), 31-33.
19. Anderson, A., Huttenlocher, D., Kleinberg, J. et Leskovec, J. (2014). Engaging with massive online courses. Dans C.-W. Chung (dir.), *Proceedings of the 23rd international conference on World Wide Web* (p. 687-698). doi:10.1145/2566486.2568042
20. Gianluca G. (2014). *Social Media Technologies in Distance Learning: New Capabilities for Future Higher Education*. International conference, new perspectives in science education (3rd edition)