Mapping student engagement in a massive open online course (MOOC)

Sterling Quinn, Anthony C. Robinson

Department of Geography, GeoVISTA Center, and John A. Dutton e-Education Institute, The Pennsylvania State University, 302 Walker Building, University Park, PA 16802, USA

Abstract. The massive open online course (MOOC) has appeared as an alternative form of distance instruction designed to scale to global audiences. Although a fair amount is known regarding who takes MOOCs, we know much less about the geography of MOOC student engagement beyond survey results that summarize students' home countries. This paper characterizes spatial patterns of learner engagement in MOOCs by mapping students' locations, attributes, and learning behaviors from a cartography MOOC offered by Penn State University titled Maps and the Geospatial Revolution. We associate course participation statistics with student locations (geocoded using IP addresses) and summarize the results using choropleth maps featuring hexagonal bins. The resulting maps show rural and urban differences in participation across countries and continents at a level of granularity not previously depicted in the MOOC literature. Themes we explore include course enrollment, course completion, engagement with course forums, gender distribution, and percentage of students "dropping out" of the course after the first exam. These findings can help online educators to better understand student behaviors in massive, global courses.

Keywords: MOOC, distance education, online education, binning

1. Introduction

Massive open online courses (MOOCs) are a relatively new form of distance instruction offered by institutions of higher education (Daniel 2012). Thousands of students at a time may participate in the courses, for reasons as diverse as gaining new job skills, challenging oneself intellectually, seeking entertainment, or simply looking for stimulating social interaction (Breslow et al. 2013). The web-based nature of the courses makes them accessible to

students throughout the globe, and a single offering of MOOC typically involves tens of thousands of students coming from dozens of countries.

Initial speculations that MOOCs would represent a breakthrough education mechanism for reaching the impoverished have been tempered by data showing that most MOOC participants already have levels of education that surpass the norms for their countries, especially in the developing world. These results were reported by Christensen et al. (2013), who also found users of the popular MOOC platform Coursera to be typically young, male, and employed, with the majority hailing from developed countries.

Although we know basic information about who takes MOOCs, very little mapping has been attempted with MOOC participant data beyond showing enrolled students as a dot on a world map or summarizing the number of enrollments at the country level. There are several major limitations with these approaches. First, levels of connectivity and access can vary greatly within a single country, affecting the way that a students in different locations approach a MOOC (Liyanagunawardena et al. 2013), therefore the country is not an adequate scale for understanding all aspects of MOOC participation. Efforts to map each student as a dot may provide some inkling of rural representation in the MOOC, but also run the risk of underemphasizing the size of the urban majority where many dots are placed on top of each other.

Additionally, it seems reasonable to believe that cultural differences such as those identified by Hofstede et al. (2010) might explain geographic variance in engagement with course components such as forums, quizzes, and peer assessments. Maps of MOOC participation can go beyond simple enrollment levels to show geographies of student activity within the course. An analysis of these maps could help instructors exhibit greater sensitivity, understanding, and preparation for global variation of student behavior in their courses.

In this paper we begin characterizing and exploring spatial patterns in MOOC learner behavior by studying a geography-themed MOOC entitled *Maps and the Geospatial Revolution* (hereafter referred to as the Maps MOOC). Offered by Penn State University for the first time in 2013 on the Coursera platform, this MOOC has attracted over 75,000 enrollments to date from over 200 countries. The course is a 5-week introductory class on Cartography and GIScience. As one of the first geography-focused MOOCs, it attracted over 48,000 students in its first offering, with over 3,000 completing the course with a passing grade (a seemingly low, but actually common rate of completion for a MOOC). Details about the development and delivery of this course are described in Robinson et al. (2015).

After summarizing related work in understanding the geography of MOOC participants, we will describe how we mapped and interpreted the geographic variation in student participation in the Maps MOOC, such as course enrollment, course completion, engagement with course forums, gender distribution, and the percentage of students "dropping out" of the course after the first exam. These findings can help learning designers and educators better understand MOOC student behaviors that appear in the massive global classroom.

2. Related work

Maps of MOOC participants have appeared in scholarly works summarizing the pedagogy of specific MOOCs and the demographics of their participants. For example, in a paper describing the Massachusetts Institute of Technology's (MIT's) first MOOC "Circuits and Electronics", DeBoer et al. (2013) present a map where each participant in the course is represented as a dot geocoded from his or her IP address. Dots are solid across Europe and the east and west coasts of the United States. India and some portions of South America also show many participants. Elsewhere participation is sparse and seems to be collected in urban areas. A unique aspect of this particular work is a table of the top countries for enrollments that includes information about the number of students completing the course, the mean number of points received per student, and the mean amount of time spent on homework. Some countries, such as Spain, exhibit markedly high values for all these variables, and while it would seem helpful to map these results, no maps are supplied.

Christensen et al. (2013) surveyed students in all of the University of Pennsylvania's MOOCs, asking them to identify their country of origin. They divided their analysis into several categories of economic development, including United States, Non-US OECD (Organisation for Economic Co-operation and Development) countries, BRICS (Brazil, Russian Federation, India, China, South Africa) countries, and other developing countries. Although no maps are provided, the detailed tables in this work are helpful toward revealing the gap that exists in developing countries between the education levels of the general population and residents who take MOOCs.

Related to our focus in this paper, Robinson et al. (2015) describe an initial approach at understanding the geographic distribution of participants in the Maps MOOC. During the first week of the course, students were asked to voluntarily place a "pin" of themselves on a web-based map of the world. The authors then took this information and created a map binned to 100km

hexagons, which still appear nearly as small as dots at a worldwide scale but at least could be classified and symbolized based on the number of included students. The resulting distribution is similar in many ways to that observed in the MIT circuitry course described above.

3. Methods

The Maps MOOC is hosted on the Coursera platform. As provisioned by the Coursera platform's terms of use, data was collected throughout the course in the form of discussion forum posts, quiz scores, grades from assessments, and basic demographic information from surveys of enrolled students. The last known IP address of the student was also collected. Christensen et al. (2013) found little variation between students' self-reported home countries and their countries as determined by Coursera-supplied IP addresses.

We assigned latitude and longitude coordinates to the IP addresses using a commercially available geocoding web service from MaxMind and were able to locate 35,783 of 49,000 students in this way from the data collected during the first session of the Maps MOOC in Summer of 2013. Out of these, 5,626 students had also completed a Coursera survey asking detailed demographic information such as gender and birth year, which we joined to the main table of students and used to create some of the maps shown below.

To derive the measures reported in our maps, we wrote scripts and custom queries to summarize the data in ways not provided "out of the box" by Coursera. For example, we wrote a Python script to tally the number of posts on the discussion forums initiated by each student. Our analysis ultimately used two datasets that we loaded into a PostgreSQL database using the PostGIS extension for spatial data. The first was a point dataset of individual students containing each student's location and attribute information. The second was a polygon dataset of tessellated hexagon-shaped bins, with hexagons measuring 2 degrees across from straight edge to straight edge. The bins were spatially joined to the points such that each point received an attribute representing the ID of its corresponding bin.

To make our maps, we applied structured query language (SQL) WHERE clauses to the student points and used the GROUP BY function to count the number of students falling in each bin. The counts were then written as attributes onto each hexagon and symbolized in a choropleth map. We used a sequential color ramp derived from ColorBrewer (Harrower & Brewer 2003). Attributes were typically classified using a quantile scheme of four classes, with class breaks rounded up to the nearest whole number or per-

centage point; however, we used the same class breaks in all the enrollment maps (ie, Figures 1 – 3) in order to more easily compare patterns. Finally, when working with percentages we excluded bins with relatively few students, thereby removing some noise from the spatial patterns that appear in the maps.

4. Results

The raw count of enrollments (Figure 1) shows that Maps and the Geospatial Revolution attracted a global audience, although some parts of the globe are more represented than others. The patterns in this map tend to hold with the set of top countries observed in MOOCs by DeBoer et al. (2013) and Christensen et al. (2013): the US on top, followed by India, Spain, Canada, and the UK. The advantage of this map is that the distributions of enrollments within the bounds of each of these top countries can now be more clearly seen. For example, enrollment rates are far from uniform within India.

Figure 1. Number of student enrollments in the Maps MOOC.

Most MOOCs see a large number of initial enrollments, with only a subset of these moving on to actually interact with the course. Figure 2 maps the number of students who received a grade in the Maps MOOC, meaning that they completed at least one of the weekly quizzes or the Final Exam. The same class breaks are used as Figure 1, and a close comparison between these two maps show some attrition. Values are generally lower, and the value of the highest bin has shrunk from 1128 to 344.

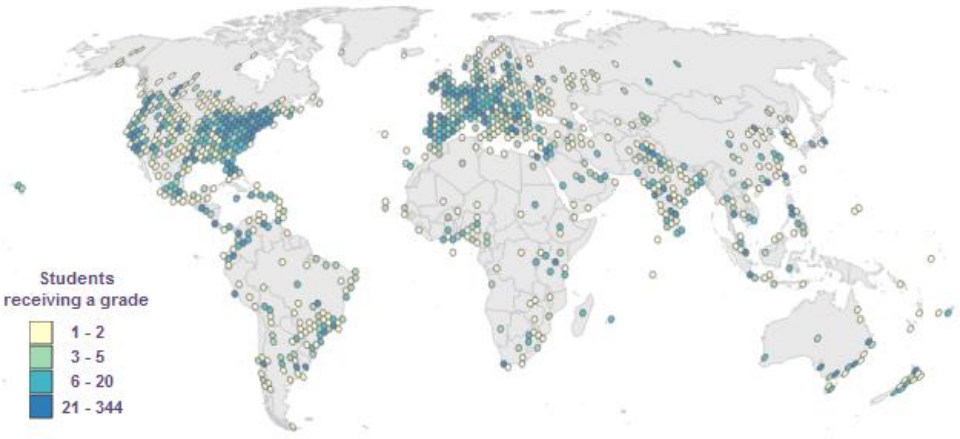

Figure 2. Number of students receiving a grade on at least one assignment.

A more drastic reduction is seen when viewing the number of students who received a grade of 70% or higher and thereby earned a passing grade in the course (Figure 3). The prominence of the Iberian peninsula on this map seems somewhat surprising at first, but corresponds with the high activity rates among Spanish MOOC participants reported by DeBoer et al. (2013). Further work to compare the difference of bin values between these maps (or normalize by the number originally enrolled) might indicate which regions of the world suffer most from losses of students between the stages of enrollment, engagement, and course completion.

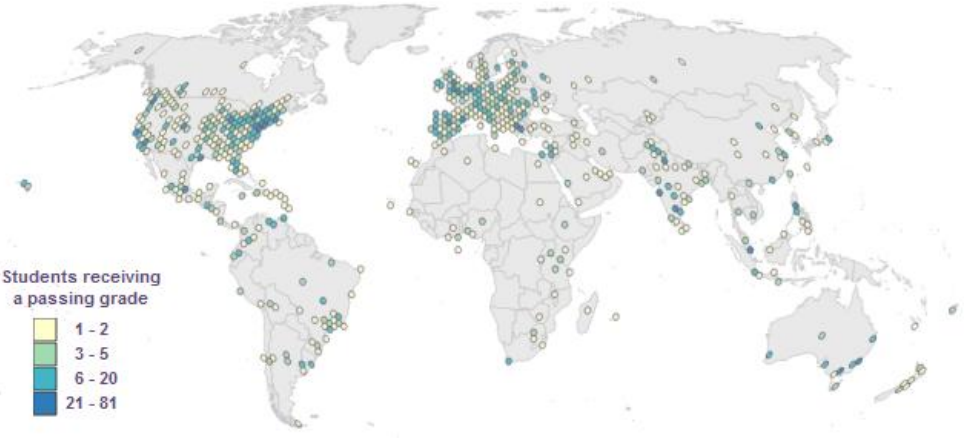

Figure 3. Number of students receiving a passing grade in the course.

The demographic information supplied by a subset of the students offers further information about who is enrolling in the course. Participants are asked to identify their gender, year of birth, country of birth, employment level, proficiency in English, and other metrics. Unfortunately, the limited number of respondents makes it difficult to make inferences about the spatial nature of some of these variables outside highly populated areas; however, Figure 4 shows trends in the gender breakdown of students when bins with 10 or more survey respondents are shown. This map reveals low percentages of women students across most of the Global South and parts of Europe, whereas most of the bins with higher percentages of women fall within the United States.

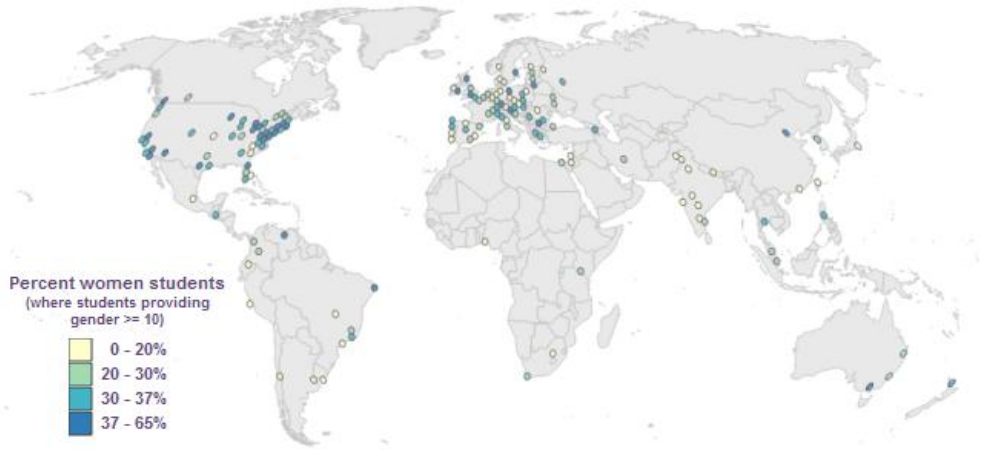

Figure 4. Percent of students who reported their gender as female.

Course engagement statistics, on the other hand, can be derived for all students, allowing a higher filter threshold on the maps than was possible when using the Coursera demographic survey data. For example, we tallied the number of students having at least one interaction with the course discussion forums and mapped this in Figure 5, showing only bins where the number of students was greater or equal to 25. Participation appears highest in the United States, and drops off drastically when one ventures outside of English and Spanish speaking countries (the top two languages spoken by participants of this MOOC). However, there are exceptions, and we speculate that the patterns in this map could be influenced by cultural differences as much as language differences.

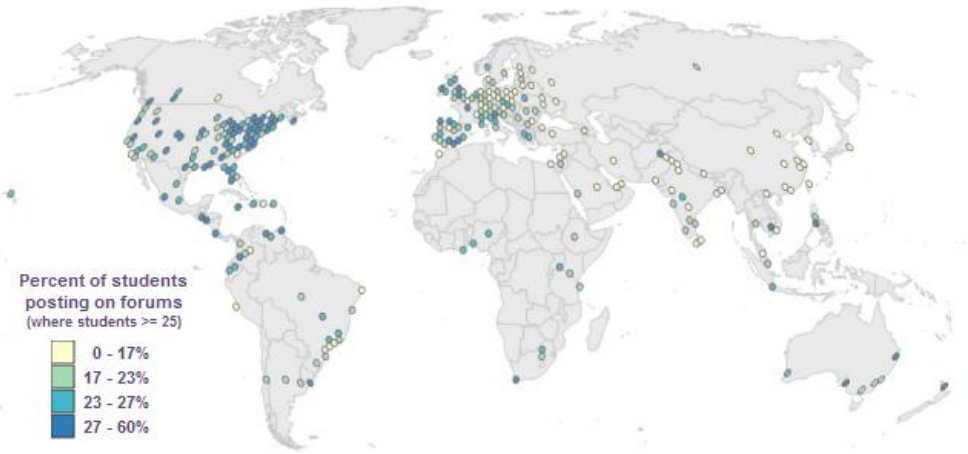

Figure 5. Percent of students who posted at least once on the discussion forums.

MOOCs tend to suffer from an attrition phenomenon noted earlier in this paper), wherein students are "lost" between each assessment. Breslow et al. (2013) noted the same trend we saw: that the number of students lost is greatest during the first few weeks. To further investigate the notable drop between students completing Quiz 1 and the rest of the quizzes in our MOOC, we mapped the percentage of students who "dropped out" after Quiz 1 (Figure 6). To reduce noise in the patterns, only bins with at least 25 students originally completing Quiz 1 are shown.

The United States showed the highest dropout rates, with Europeans being much less likely to drop out. The cluster of low dropout rates in southeastern Europe is particularly notable. Christensen et al. (2013) found that in the United States, topic curiosity (rather than seeking job skills) is a bigger motivation for taking a MOOC than elsewhere in the world, and this may explain higher dropout rates.

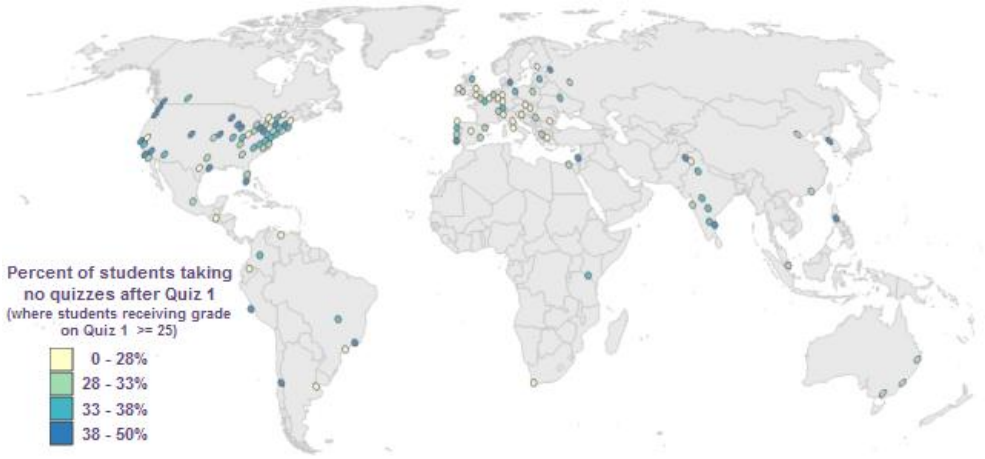

Figure 6. Percent of students who "dropped out" after Quiz 1.

Another measure of student engagement and persistence is the tendency to retake quizzes in order to achieve the highest possible score. The Maps MOOC offered four quizzes and a single final exam. The quizzes could be attempted for a grade up to three times, with the most recent score taken as the official grade for the quiz. The final exam allowed only one attempt at a grade. Numerous students attempted the quizzes multiple times in order to improve their grades.

To identify those places where students were most intent on improving their grade, we mapped the percentage of students attempting all four of the quizzes more than once (Figure 7). We acknowledge that this is a less-than-perfect measure of persistence because a student achieving a perfect score on any one of the quizzes presumably would not appear in the results; however, it is one clear measures of persistence in the course that we can quantify.

Bins where 25 or more students received a grade in the course were included on this map. Geographic patterns in this map are harder to discern, but in general there are lower percentages in the United States and Western Europe, with some areas of Asia and southeastern Europe showing high rates of persistence in retaking quizzes.

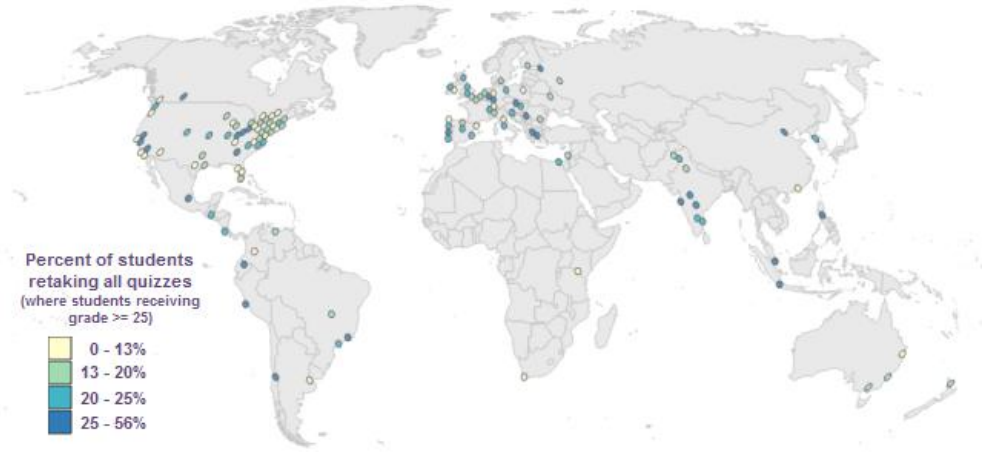

Figure 7. Percent of students who took all quizzes more than once.

5. Summary and conclusions

The Coursera MOOC platform provides enough data to map and identify trends in both enrollment and participation habits; however, the data must undergo some manipulation prior to being mapped. Choropleth maps with tesselated bins can show trends at a finer-grained level than a country, without causing confusion from multiple overlapping symbols. When percentages are involved (rather than raw counts), it can still be difficult to discern patterns in rural areas. The small size of the subset of students completing the demographic survey also causes challenges for interpreting trends such as birth country in a MOOC of this size. Nevertheless, we were still able to use the demographic survey results to identify a dearth of female participants in India and other areas of the Global South.

Spatial patterns were also apparent in the maps of course engagement, including discussion forum participation, dropout rate, and persistence in retaking all quizzes. Other variables that could be mapped might include the percent of students receiving at least one perfect score on a quiz, the average number of forum posts per student, and the average score of each student. Looking forward, we are initiating the design of an interactive web map interface where the user can adjust parameters on the fly to flexibly create maps of MOOC learner activity.

Our results are intended to provide useful geographical context to learning designers and MOOC instructors who want to know more about how to meet the needs of a global student cohort. For example, based on the maps we show highlighting low forum posts in the Middle East and Asia, an in-

structor might decide to create dedicated discussion forums in popular languages such as Chinese or Arabic. In general, we see a bright future ahead for the broader integration of mapping and spatial analysis to inform the design of online learning experiences by helping educators to understanding their audiences' locations, needs, and habits.

Acknowledgements

This research was funded by a Penn State Center for Online Innovation in Learning (COIL) Research Initiation Grant. The authors would like to thank Scott Pezanowski for technical assistance with this project.

References

Breslow L, Pritchard D E, DeBoer J, Stump G S, Ho A D, Seaton, D T (2013) Studying Learning in the Worldwide Classroom: Research into edX's First MOOC. Research & Practice in Assessment, 8(1), 13–25

Christensen G, Steinmetz A, Alcorn B, Bennett A, Woods D, Emanuel E J (2013) The MOOC Phenomenon: Who Takes Massive Open Online Courses and Why? Available at SSRN 2350964. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2350964. Accessed 10 March 2015

Daniel J (2012) Making Sense of MOOCs: Musings in a Maze of Myth, Paradox and Possibility. Journal of Interactive Media in Education, 2012(3), Art–18

DeBoer J, Stump G S, Seaton D, Breslow L (2013) Diversity in MOOC Students' Backgrounds and Behaviors in Relationship to Performance in 6.002 x. In Proceedings of the Sixth Learning International Networks Consortium Conference. http://tll.mit.edu/sites/default/files/library/LINC%20%2713.pdf. Accessed 10 March 2015

Harrower M, Brewer C A (2003) ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. The Cartographic Journal, 40(1), 27–37

Hofstede G, Hofstede G J, Minkov M (2010) Cultures and Organizations: Software of the Mind, Third Edition (3 edition). McGraw Hill, New York

Liyanagunawardena T, Williams S, Adams A (2013) The Impact and Reach of MOOCs: A Developing Countries' Perspective. eLearning Papers, (33). http://centaur.reading.ac.uk/32452/. Accessed 10 March 2015

Robinson A C, Kerski J, Long E C, Luo H, DiBiase D, Lee A (2015) Maps and the Geospatial Revolution: Teaching a Massive Open Online Course (MOOC) in Geography. Journal of Geography in Higher Education, 39(1), 65-82