



Contents lists available at ScienceDirect

Computers & Education

journal homepage: <http://www.elsevier.com/locate/compedu>

Differences in forum communication of residents and visitors in MOOCs

Oleksandra Poquet^{a,*}, Jelena Jovanovic^b, Shane Dawson^c^a University of South Australia, Centre for Change and Complexity in Learning, BH3-21, City West Campus, Adelaide, 5000, South Australia, Australia^b University of Belgrade, FOS - Faculty of Organizational Sciences (School of Business Administration), Department of Software Engineering, Jove Ilica 154, 11000, Belgrade, Serbia^c University of South Australia, Education Futures, Level 1 De Lissa Building, Magill Campus, South Australia, Australia

ARTICLE INFO

Keywords:

Massive open online courses
Online forums
Communication
Learner profiles
Networked learning

ABSTRACT

Facilitation of social interactions in Massive Open Online Courses can benefit from conceptualizing forum sub-populations through a networked lens. Such a lens implies that forum interactions represent a network of learners with heterogeneous levels of commitment to forum activity. A networked lens helps capture forum communities, so-called *forum residents*, as well as the rest of the crowd, also known as *forum visitors*. The current study empirically demonstrates the differences between these forum sub-populations. Building on a large dataset of manually labelled discussion threads in four edX MOOCs, our findings uncover two distinct patterns in the communication of posters with different length of forum commitment. As the courses progressed, all types of communication decreased for forum visitors, but both socio-cognitive and informational queries increased for forum residents. We find that the communication dynamics of committed forum posters cannot be observed when the entire forum population is examined in its entirety. Further, the study profiles learners around discussion type sequences. We show that communication topics of forum visitors appear narrow and topical when compared to the diversity of topics by resident posters. This paper offers a foundation towards the scaling of social teaching practices for personalised learning of both residents and visitors as well as community development in massive open online courses.

1. Introduction

Since their inception in 2008, Massive Open Online Courses (MOOCs) have occupied their place among online education models. Despite the widespread adoption, research suggests that teaching practices adopted in MOOCs have been limited (Zawacki-Richter, Bozkurt, Alturki, & Aldraiweesh, 2018). Littlejohn and Hood (2018) evaluated the state of MOOC design and instruction as not forward-looking. In part the limitations stem from thinking of MOOCs as formal online courses, such as those delivered at the university for credentialing (Joksimovic et al., 2017). This reasoning is also observed in the effort MOOC instructors apply to support learners to complete the course and engage with all course assignments. In contrast to these efforts, research shows that such intentions do not represent most MOOC participants (Kizilcec & Schneider, 2015). However, MOOC research has also oversimplified learner

* Corresponding author.

E-mail address: sasha.poquet@unisa.edu.au (O. Poquet).

related factors, focusing on profiles of learners likely to complete (Deng, Benckendorff, & Gannaway, 2019). Supporting learners whose goals diverge from course completion is equally important to advance the vision for open online learning: to encourage learners to follow their own goals, rather than conform to what is desired by the instructor (Littlejohn & Hood, 2018).

The same critique applies to the facilitation of online interactions in MOOCs, which is a main strand in MOOC research (Zawacki-Richter et al., 2018). Facilitation and analysis of social interactions in MOOCs – the focus of this study – is also often approached through the lens of formal online education. MOOC forums are predominantly examined in relation to retention and success with quizzes (Almatrafi & Johri, 2018), ignoring the diversity of motivations for MOOC participation. Importantly, analytical frameworks applied to MOOCs are transferred directly from traditional theories developed to understand formal online courses with bounded small-size cohorts. These formal theories carry socio-constructivist assumptions of forums being bounded groups that accumulate interactions over time. However, as shown by research (Poquet et al., 2018; Kovanovic et al., 2019), MOOCs are not suited for the direct transfer of such socio-constructivist theories.

Instead, facilitation of social interactions in open online education should be developed with the notion that MOOC forums are networked environments (cf. Joksimovic et al., 2017), not groups. Conceptually speaking, viewing online interactions as a network has been juxtaposed to conventional socio-constructivist views (Jones & Esnault, 2004). A networked view implies the presence of both a group with regular repeated interactions, as well as the presence of actors who are peripheral and intermittent in their interactions. Networked learning emerged as a socio-technical perspective on digital learning, an alternative to more stringent socio-constructivist theories of online learning. Analytically, networked learning affords a view that while some learners connect through more sustained interactions as they would in a university course, for others intermittent exchanges suffice (Haythornthwaite, 2019; Jones, 2004; Littlejohn & Hood, 2018).

A view of MOOC forums as networks, i.e. collectives that are broader than bounded communities, has been validated by early empirical research in MOOCs. Gillani and Eynon (2014) suggested that forums garner the presence of both crowds and communities. However, later MOOC research maintained the bounded-group lens on the forums (e.g. Moore, Oliver, & Wang, 2019) and does not frame examination of MOOC participants through the networked perspective (e.g. Boroujeni & Dillenbourg, 2018; Luna, Fardoun, Padillo, Romero, & Ventura, 2019).

This study analyses MOOC forums through a networked perspective, where learners are differentiated based on their commitment to the social activity, some engaging in community, others engaging on the periphery and intermittently (Honeychurch, Bozkurt, Singh, & Koutropoulos, 2017). To this end, the current study examines the differences in communication across sub-populations of forum participants, using the discussions by 7295 posters in four EdX MOOCs. Forum communication is characterised as content-task, content-non-task, social-task, and social non-task discussions, as well as informational administrative queries. Using negative binomial mixed effect models, we model the change in learner communication at the beginning, in the middle, and in the end of the course, as well as control for the differences in regularity of posting activity. Learners with different posting regularity and commitment to forum use here are referred to as *residents* and *visitors* (White & Le Cornu, 2011). Using sequence analysis of discussion thread types at a learner level, we profile forum participants through both their discourse and behaviour. The findings demonstrate differences in patterns of activity and discourse, paving way forward to teaching practices that embrace learner diversity.

2. Research framework

2.1. Visitors and residents in networked spaces

Socio-technical and networked theoretical premises are implicit within this study's view of digital learning. MOOC forums can be viewed as networked spaces where open online course participants communicate, driven by different purposes, interests, and expectations. Conceptions of networked learning are based on the socio-technical view of learning in digital settings. Jones (2015) and Jones & Esnault (2004) juxtaposed the context of learning in a network with socio-constructivist understandings of a community. The notion of a community in educational theories is built on the premise that knowledge is constructed through shared interactions; when framed within such theories, the social context of knowledge building is that of an evolving bond-based community. This implicit understanding of a community in online learning presumes continuity of interactions within a bounded group of learners, that with time, results in strong relations.

According to Jones (2004), a networked context, as often observed in digital learning, does not privilege the strong relationships that imply closeness and unity of purpose as much as a *community* in a formal course would. Networked learning as a lens therefore allows for the inclusion and consideration of relationships that differ in strength; that is, both closely-knit and loosely coupled groups. This view of digital learning as that of a network, rather than a community is resonant with Haythornthwaite and de Laat (2012) who suggested that networked learning emerges through varying quality of tie content underpinning learning networks. Connectivism is another socio-technical perspective, that focuses on the structure of communication (Siemens, 2013) by introducing a learning infrastructure that includes structurally central and peripheral actors.

Given the exploratory nature of the present study, we adopt White and Le Cornu's (2011) framework of online engagement in digital spaces to operationalize commitment to social learning. This framework aligns with the theoretical premises of networked learning as it lends itself to positioning MOOC forum use through a lens of commitment that is reflective of learner intentions. For instance, an individual may choose to *visit* a course and discussion activities as necessary, or to *reside* and make the digital space one's own. Individuals indicative of a visiting attitude are likely to have a pre-defined goal or a task. Such learners select an online resource or activity that meets their perceived needs. Once the task is completed, and requirements met, the individual leaves the digital space. In contrast, a resident's attitude implies a stronger motivation "to connect to, or to be with, other people" (White & Le Cornu, 2011,

sec.IV.2).

This study draws on White and Le Cornu's framework to investigate discrete sub-populations of MOOC forum participants. We examine if the differences inherent within the forum visitors and resident metaphor can be captured through the observed posting behaviour as well as the discussion topics learners engage in. We assume that posters with more pronounced resident behaviour demonstrate a higher volume of activity and persistence. Correspondingly, we approach visitor behaviour as that of a lower volume of forum activity and lower level of interest in forum engagement. Finally, we hypothesise that diversity and quantity of content topics posted by visitors and residents differ. We suggest that topic-wise, visitors have more consistent discourse engagement. That is, engagement reflective of and aligned with their specific goals. In contrast, residents engage in more diverse discussions, driven by their interest to connect through social learning. The present study examines the validity of these propositions.

2.2. MOOC forums: what research tells us

Studies of learner activity in MOOCs demonstrate that learners follow diverse engagement patterns (Coffrin, Corrin, de Barba, & Kennedy, 2014; Ferguson & Clow, 2015b; Hill, 2013; Kizilcec, Piech, & Schneider, 2013; Kovanović et al., 2019). Driven by different learning goals and motivations, learners exhibit selective preferences in engaging with course resources (Kizilcec & Schneider, 2015). Eynon et al. (Eynon, Gillani, Hjorth, & Yasserli, 2014) noted that just-in-time learners sample only a subset of the available videos, whereas students seeking certification focus on assessment activities. This trend of selective preference is particularly exacerbated in discussion activities. Only a fraction of learners participate in course forums, and those that do, present vastly different temporal patterns of activity (Hecking, Chounta, & Hoppe, 2016; Shirvani Boroujeni, Hecking, Hoppe, & Dillenbourg, 2017; Tang, Xing, & Pei, 2018; Yang, Wen, Kumar, Xing, & Rosé, 2014). Some learners join forums for short bursts of activity, without returning to them, while others show more persistent and regular use throughout the duration of the course (Poquet, 2017; Gillani & Eynon, 2014; Hecking et al., 2016; Jiang, Zhang, Liu, & Li, 2015).

Research on MOOC discussions clearly illustrates that participation is variable in terms of quantity, frequency, persistence, and commitment (Baker, Evans, Greenberg, & Dee, 2014; Coffrin et al., 2014; Ferguson & Clow, 2015a; Khalil & Ebner, 2017; Kovanović et al., 2019). These differences can be captured through White and Le Cornu's metaphor of 'visitor' and 'resident' behaviour (2011). Lurking is another typical activity for a MOOC learner. Lurking refers to legitimate peripheral participation in the network of learners (Honeychurch et al., 2017). This behaviour describes learner preference to engage with the content exclusively through reading and listening, rather than creating new content (Sun, Rau, & Ma, 2014). In MOOCs most people read the forums while observing the others interact (Anderson, Huttenlocher, Kleinberg, & Leskovic, 2014; Seaton, Bergner, Chuang, Mitros, & Pritchard, 2014), for reasons ranging from the lack of time to the lack of interest to engage (Bozkurt, Koutropoulos, Singh, & Honeychurch, 2020). Since lurkers do not engage in explicit participation, they are not captured by the resident-visitor metaphor and are not the focus of this study.

Despite observed differences in temporal and activity-level, research on inter-group comparisons of forum posters has been limited, in particular, research involving the analysis of content that learners post online. To date, research related to the content or topics of forum postings has been carried out at the forum-cohort level, not at the level of learner sub-populations. The analysis of the entire forum contributions illustrates that both quantity and diversity of topics decreases as the course progresses (Ramesh & Getoor, 2018). However, Poquet & Dawson (2016) show that when the posting activity of the more persistent learners is examined separately, the amount of curriculum-related discussion within this sub-population actually grows in quantity throughout the course. This suggests that examining cohort-level activity without an explicit consideration of time as a factor can mask insights available into the forum dynamics that occur at the sub-cohort level.

2.3. Beyond on-task focus and cohort-level analysis of MOOC forums

Analysis of forum posts in relation to learning and course communication has been mostly conducted at the cohort-level, with a focus on automated analysis of discourse or on the content analysis of student-generated text (Arguello & Shaffer, 2015; Ezen-Can, Grafsgaard, Lester, & Boyer, 2015; Kovanović et al., 2016; Rossi & Gnawali, 2014; Wise & Cui, 2018). Incorporating the entire student population is useful for studies that automate feedback in forums, to rapidly 'inform' or 'sort' questions and support requests (Atapattu & Falkner, 2016; Chandrasekaran, Ragupathi, Kan, & Tan, 2015; La Vista, Falkner, & Szabo, 2017), but it does not offer insights about the dynamics of communication as reflective of learning.

What we currently know about the dynamics of communication in networked spaces at scale is limited. Much of the insight is derived from approaching MOOCs as equivalent to forums in formal online courses. However, MOOC forums do not reflect the same dynamics. MOOC posters range significantly in their level of commitment, motivation and underlying interests in the topic. These differences manifest in the varied levels of participation and length of engagement in the course. The diversity of participation quantity and frequency results in a lack of group boundaries and lower levels of shared experiences among individuals. The presence of a shared experience is an important element in developing a shared sense of community. Despite these differences, analysis of MOOC forums is often conducted at the forum-cohort level. Furthermore, interpretation of these findings is applied through theories of learning established in formal education – a context that differs to and pre-dates MOOCs.

Generally, social learning theories do not transfer well to open learning environments such as MOOCs. For example, socio-constructive models that underpin many current online learning practices, were built on data collected in small groups of learners taking formal university courses. The education context for these learners results in numerous opportunities to build a shared history due to repeated courses and longer-term communications. As such learners in this context regularly interacting with peers are more likely to engage in knowledge construction. In MOOCs, the focus on knowledge construction (observed in content on-task topics) is

meaningful only in relation to those learners who have repeated interactions.

Given the misalignment between the MOOC context and socio-constructivist premises, interpreting current empirical studies is a challenge. Overall, analysis of different topics in the MOOC forums at the forum-cohort level shows a decrease of volume as the course progresses (Kellogg, Booth, & Oliver, 2014; Ramesh & Getoor, 2018; Rossi & Gnawali, 2014; Shirvani Boroujeni et al., 2017). On the other hand, analysis of content posted within a sub-population of regular posters, conceptually comparable to a group in a formal course (Poquet & Dawson, 2016), revealed the growth of curriculum-related discussions in volume. A comparison between the different poster sub-population is required for meaningful interpretation of these findings.

A further limitation of the current approaches to content analysis in MOOC forums is their dominant focus on content-related task-oriented discussions (Cheung, 2014; Hecking, Chounta, & Hoppe, 2017; Ramesh & Getoor, 2018; Stump, DeBoer, Whittinghill, & Breslow, 2013; Wang, Wen, & Rosé, 2016; Wang, Yang, Wen, Koedinger, & Rosé, 2015; Wise, Cui, Jin, & Vytasek, 2017; Wise, Cui, & Vytasek, 2016). However, examination of off-task communication is also required to identify precursors to deeper knowledge building processes and post for stronger group cohesion. From a socio-constructivist perspective, while knowledge construction is tightly linked to cognitive and on-task communication across repeated learner interactions off-task communications can also promote and encourage group cohesion and therefore, knowledge construction. For example, the exchange of “interpersonal” messages, not related to formal content of subject matter’ (Henri, 1992, p. 126) provide more effective “impersonal” task-oriented communication. That is, discussions related to socio-emotional processes are necessary for a deepening of group cognitive engagement (Akyol & Garrison, 2008). In the MOOC setting, most learners do not engage in socio-emotional discussions due to their short-term visits and often differing motivations and learning needs. The fleeting nature of these exchanges does not aid the development of discussions for knowledge construction. The investigation of off-task communication in MOOCs has thus far has been limited.

With its specific research focus, the current study seeks to extend research beyond the analysis of content-related task-oriented posts at a cohort level, to further inform our understanding of MOOC forums as spaces for networked learning. We statistically analyze change and sequence of manually labelled forum posts to identify distinct groups of learners along the resident – visitor continuum. Such groups are characterised by unique patterns of temporal distribution of the discussion topics (e.g. content and non-content, on-task and off-task). The distinctive patterns exhibited by the sub-groups confirm the need to treat distinct poster sub-populations differently. As communication across these sub-groups unfolds, so does the forum as a social learning space.

2.4. Research questions

To address the lack of analysis of the differences in MOOC forum communication (e.g., content and non-content, on-task and off-task) and to examine the changes in posting behaviour across different learner sub-populations, this study poses the following research questions:

RQ1: How do topics of communication in MOOC forums change throughout the course and across posters with different forum activity?

RQ2: What profiles of MOOC forum posters can be identified through the sequence analysis of the discussion topics throughout the course?

In addressing these two research questions we can extrapolate on the evolution of MOOC forum as a networked space for both group/community and individual learning.

3. Methods

3.1. Overview of methods

The study examines forum activity in four xMOOCs (two – in engineering, one – in computer science, one – in data analysis). These edX courses were delivered in 2013–2014. The instructional design of the analysed courses was typical for xMOOCs, with a high volume of content in the forms of videos, weekly delivered quizzes and assignments, and minimal targeted activities to promote social learning in the forums. All MOOCs had high enrolment numbers ($M = 40878$, $SD = 10972$). Three of the courses ran for 8 weeks; the fourth course ran for 11 weeks.

This quantitative study analyses forum communication change using mixed effects binomial regression. The study also profiles poster discussion types through sequence analysis. Both regression and sequence analyses used discussion threads previously labelled through a *qualitative content analysis*, i.e. not automated but conducted manually (Krippendorff, 2004, p. 102). Current paper only reports quantitative analyses that use labels from qualitative content analysis. The method section describes coding categories used for discussion labelling since understanding them is integral to understanding the results. The process of content analysis and its results are detailed elsewhere (Poquet, 2017, Chapter 5).

To address the first research question, we modelled the change in counts of different discussion topics using binomial negative mixed effect regression. To show differences in communication at distinct times in the course, the course was divided into discrete time-slices using the design of the courses and its activity (Poquet, 2017). First two weeks were aggregated into the first time slice, referred to as ‘the beginning of the course’, the last two weeks were aggregated into the third time slice, i.e. ‘the end of the course’. In each of these periods, posts within each thread-level-topic (identified through content analysis) were counted at the level of individual posters and used as dependent variables in the regression models (Section 4.4). We controlled for the sub-population differences in the

regression models, by including the number of weeks a learner posted and the number of weeks between the first and the last week of posting. To control for the forum facilitation differences, fixed effects related to course design and underlying forum dynamics were also included.

To address the second research question, we created a sequence of discussion topics at the learner level, i.e. each learner post was labelled based on the discussion thread it was situated within. These sequences were then clustered based on the similarity of temporal distribution of topics for each participant. The volume and duration of forum participation were used to validate each profile.

3.2. Content analysis of forum threads

Quantitative analysis of forum discussions incorporated labels derived from qualitative content analysis conducted in another study. The labels were assigned as a part of a separate effort, in relation to *the intentions learners had when posting* rather than *what the post was about*. Evaluation of the *quality* of the posts was beyond this study's scope. The label was assigned to the discussion thread, and each post within the thread was used as the unit of analysis, taking on thread label. This was done to capture the different volume of communication types. The dataset modelled in this study comprised a total of some 4260 discussion threads, i.e. 16195 posts, across the four MOOCs ($M_{\text{coded_threads_per_course}} = 1065$, $SD = 123$; $M_{\text{coded_posts_per_course}} = 4048$, $SD = 253$).

Discussion types used in this study can help interpret whether or not the posters were talking to others to understand the curriculum, were reflecting on the self, or making sense of the educational environment. Orthogonal to these broader intentions, discussions were coded as on-task or off-task communication. Mixed and meta-cognitive discussion types, both present in the labels, were excluded from modelling due to scarcity. [Tables 1a and b](#) offer an overview of the coding scheme and example posts for each category. [Table 1c](#) explains how posts and discussions were transformed into units of analysis detailed in [Section 3.3](#).

Curriculum-Related Discussion Threads. The curriculum-related topics capture posts where the learner's intention was to inquire about the course's curriculum, such as seeking clarification of the course concepts, lectures, assignments, and implementation of the subject matter in practice. Two topics were related to the curriculum: content-task (CT); and content non-task (CN). The CT topic is assigned to discussion threads triggered by the content of a specific task, e.g. learner's confusion about how to solve a graded assignment, or a graded tutorial. Discussions triggered by an inquiry into the course materials or concepts were labelled as CN (e.g. clarification of lectures, correction of course material mistakes, interest-based discussion of out-of-the-course discoveries).

Setting-Related Discussion Threads. In setting-related discussion threads the learners inquired about the educational setting of a MOOC, trying to understand how the setting *works*. Two topics fall into this category: administrative and technical (INF) for closed-ended queries; and social non-task (SN) for open-ended queries.

The INF topic embraces informational queries posted about the educational setting, prompting a closed-ended response, e.g. common technical troubleshooting (i.e. IT-help desk queries), how to download course videos, how to read the progress bar, how to submit assignments. It also includes administrative questions related to course deadlines, eligibility to start the course, course requirements, and so on.

Learner sense-making of the educational setting of a MOOC was not limited to simple Q&A. A learner may have wondered why the content was being taught in a particular manner, why an assessment was designed in a certain way, for what purposes and how best to use the forums, inquiring into the value of MOOC learning in relation to the learner's career aspirations, why other learners were taking the course, and so on. Such open-ended Social Non-Task (SN) threads were oriented towards the group and were not required or necessary to complete any given task. The SN topic is associated with early meet and greet threads, but also later negotiations around organising forum communication and accepted social norms.

Self-Related Discussion Threads. A self-related discussion thread (Social Task, ST) included an emotional statement that a learner shared with peers on a curriculum task they were undertaking. Although such statements and discussions were triggered by a learning task, the intention of these messages was to share or reflect upon a personal skill in relation to the learning experience. For instance, whereas an emotionally written CT discussion might be a plea for help, the intent of an ST discussion would be an appeal to emotional support, venting, or reassurance needed when solving the task. ST threads emphasise that while cognitively engaging with the course concepts, individuals were also learning about themselves, as per the theories of self-efficacy ([Bandura, 1982](#)) or self-theories ([Dweck, 2000](#)).

3.3. Negative binomial mixed effects model

To address the first research question, binomial negative mixed-effects regression ([Agresti & Kateri, 2011](#); [Long, 1997](#)) was used to

Table 1a
Overview of discussion types in forum communication.

	1. Curriculum-Related	2. Setting-Related	3. Self-Related
a) On-Task	1a. Content-Task (CT) Discussions "How do you solve this task?"	2a. Administrative and Technical (INF) Queries "Can't submit the assignment, has the deadline passed?"	3a. Social-Task (ST) Discussions "This was the hardest tutorial I have ever done in my life."
b) Off-Task	1b. Content Non-Task (CN) Discussions "I was wondering why wind energy is not widely used in my country."	2b. Social Non-Task (SN) Discussions (negotiation, group-related) "I think we are all here to learn not to ace the content."	n/a

Table 1bAuthentic Examples of Each Discussion Types (for full Coding Framework, please see [Appendix A](#)).

Content-Task Discussion	<p>Learner A. Without any Up converter: only photons 2 eV & above are absorbed which is applicable for photons around 620 nm and below. All calculations as usual & let's assume current density Jsc1</p> <p>With Up converter 1: 2 photons (one from B, another from C) add up @ 65% success to make 1 photon with > 2eV and you get your Jsc2' but the total Jsc2 is the sum of Jsc1 & Jsc2'</p> <p>Learner B provides an extensive response.</p> <p>Learner C agrees.</p> <p>Learner D adds a short disagreement.</p> <p>Learner B addresses it.</p> <p>Learner D thanks.</p>
Content Non-Task Discussion	<p>Learner A: In the above lecture (at time 3.42 s) the prof. says that n & p regions are not electrically neutral. Why? Since doping atoms are electrically neutral (all atoms are), and they become positively charged, after adding to the base material, by giving away an electron, shouldn't the n region as a whole be electrically neutral? Same for p region. Please explain.</p> <p>Learner B explains</p>
Administrative and Technical Queries	<p>Learner A: I've done a couple of the practice questions throughout the chapter 1 and only received points for 1 question. The rest are still 0/0. I had 2 questions that I knew I got wrong (for the cost/kWh chart question that got me confused) but I had the other previous questions correct. Why did I not receive credit for them?</p> <p>TA addresses this request.</p>
Social-Task	<p>Learner A: So this first lecture came as kind of a shock to me. There was a lot of new expressions and exotic material names dropped, that I have no previous knowledge about.</p> <p>Learner B: Just as Learner A mentioned, the III-V devices are indeed photovoltaic devices made from the 3rd and 5th group of the periodic table. But don't worry, this group of solar cells (and other materials) will be discussed extensively in the upcoming weeks.</p>
Social Non-Task	<p>Learner A: Hi there. Most probably the designers of edX have thought of all the possibilities mentioned above. But remember that the courses are trying to meet higher levels of education, which mean higher commitment, attention to details and self-discipline. Please, please, read the guidelines for this course (and many others) in this link: [Collaboration Guidelines]</p> <p>Learner B: have done the MITx 6.002x course last year - the original pilot course. On that course assessed questions were allowed three attempts to get the answer and non-assessed questions allowed (seemingly) infinite attempts. Problems did arise on that course where varying the degree of precision in an answer course cause one answer to be accepted but not another. I therefore see legitimate concerns being expressed by Learner A.</p> <p>Learner C: Oh, and as a cautionary warning to anyone thinking of setting up duplicate accounts - they were spotted on the MITx 6.002x course and the offenders didn't get to complete the course. The honour code needs to be honoured!</p>

Table 1c

Transformation of Codes into units of analysis.

Discussion Thread Example	Transformation into units of analysis
<p>Learner A. Without any Up converter: only photons 2 eV & above are absorbed which is applicable for photons around 620 nm and below. All calculations as usual & let's assume current density Jsc1</p> <p>With Up converter 1: 2 photons (one from B, another from C) add up @ 65% success to make 1 photon with > 2eV and you get your Jsc2' but the total Jsc2 is the sum of Jsc1 & Jsc2'</p> <p>Learner B provides an extensive response.</p> <p>Learner C agrees.</p> <p>Learner D adds a short disagreement.</p> <p>Learner B addresses it.</p> <p>Learner D thanks.</p>	<p>The discussion is interpreted in its entirety based on its overall function. Since it is a content-task discussion, each learner in this post engaged in 1 CT posts, besides Learner B who engaged with 2 CT posts.</p> <p>In RQ1, binomial negative regression predicts the counts of the total number of posts of each type per learner at different time periods, controlling for learner activity levels.</p> <p>In RQ2, sequence analysis clusters learners on the sequence of posts they engaged in. In this example, Learner B sequence would be CT – CT, whereas every other learner's sequence would be CT. For sequence analysis sequences of all posts learner made on the forum would be clustered.</p>

model the counts of different discussion topics per individual student in three time periods within a course: the beginning, middle, and end. In particular, five negative binomial mixed effect models were fitted, each predicting the count of different discussion topics at various times in the course. The models controlled for learner sub-population differences using the total number of weeks the learner posted to the course forum, and the number of weeks between the first and the last week of posting.

To control for the forum-specific differences, we used design features identified in complementary work on the same dataset of courses (Poquet, 2017). The four forums differed with respect to the level of teacher presence and the level of community development processes, with all four forums positioned in one of the following 4 categories: 1) high teacher presence, high community development; 2) high teacher presence, low community development; 3) high community development, low teacher presence; 4) low community development, low teacher presence.

Raw counts of discussion topics (count of posts in particular discussion thread per learner) were used in modelling. Course labels were included as random effects. Fixed effects included: weeks (N) a learner posted (forum presence); weeks (N) between the first and the last week of posting (duration of interest); time in the course for which the count of discussion topics was predicted (beginning, middle or end); interaction between forum presence and time in the course; interaction between duration of interest and time in the course; elements of forum dynamics specific to the course (teacher presence; community development). The choice of phases in the courses was selected based on the exploratory analysis of activity in the course and with the consideration of course design features.

The distribution of post counts across the thread topics in four MOOC courses in the three examined time periods appeared over-

dispersed. This resulted from many learners having counts of zero for their types of communication in a particular time period. A dataset of such shape has variance higher than the mean, and Poisson or Negative Binomial regressions are deemed appropriate. We examined the goodness of fit across Poisson, negative binomial, and zero-inflated negative binomial regressions, and found binomial negative regressions to demonstrate the best fit. Two R packages were used for modelling, *pscl* (Jackman et al., 2017) and *glmmTMB* (Magnusson et al., 2017), with the same analysis replicated using both packages.

3.4. Sequence analysis

A sequence of posts based on the discussion topics (CT, CN, ST, SN, INF) was created for each student, based on the chronological order of the discussions they contributed to within a course discussion forum. The sequences were encoded in the format required by TraMineR (Gabadinho, Ritschard, Müller, & Studer, 2011), the R package used for sequence analysis. Fig. 1 provides examples of sequences in the TraMineR format. The sequences can be heterogeneous, both in terms of their length and the diversity of post topics they consist of (e.g. sequence [1] vs. sequence [4] on Fig. 1).

To cluster the sequences, the outliers were excluded. This included students with overly short sequences (i.e., sequences of length one) as well as those with overly long sequences (i.e. sequences of length above 97th percentile). The outliers, also referred to as one-timers (those with one post/thread only) and super-residents (those with post count above 97th percentile), were not used for clustering, but were the subject of exploratory analysis. Table 2 presents the number of students (sequences) and posts in the two outlier groups as well as the number of students (sequences) and posts used for clustering for each course.

To identify patterns in student posting behaviour, agglomerative hierarchical clustering, based on the Ward's algorithm, was applied on the group of sequences denoted in Table 2 as "regular cases". This clustering method has proven suitable for detecting student groups in online learning contexts (Kovanović, Gašević, Joksimović, Hatala, & Adesope, 2015). Since we were interested in identifying student profiles based on the temporal distribution of the different discussion topics they contributed to, we needed to base the clustering on a distance (similarity) measure sensitive to the distribution of different topics. According to an extensive comparative review of sequence dissimilarity measures (Studer & Ritschard, 2016), Chi-squared distance and Euclidean distance with the number of periods equal to 1 are recommended when the interest is primarily in the within-sequence state distribution over the entire examined period. We opted for the latter measure as it produced more interpretable results. To select the optimal number of clusters, dendrograms were used to identify the most plausible segmentations of the tree structure. The resulting clusters were examined through state distribution plots (present the temporal distribution of discussion topics across the overall examined period) and frequency plots (present the selected number of the most frequent distinct sequences) of the TraMineR package.

Kruskal Wallis tests followed by Mann Whitney U tests were used to compare the resulting clusters based on the forum presence (the number of weeks one posted), the duration of interest (the number of weeks between the first and the last weeks of posting), and the number of posts made. False Discovery Rate (FDR) was used as a recommended correction for preventing alpha inflation when doing multiple tests (Cramer et al., 2015).

4. Results

4.1. RQ1: change in discussions topics across poster sub-populations

RQ1 inquired how different communication topics in MOOC forums change throughout the course across posters with different forum activity (see Table 3). Five negative binomial mixed effects models, that predict that rate of change in counts for each type of discussion across four courses. The rate of posting across all discussion topics was significantly associated with forum presence, duration of poster interest, course features, and time in the course. The change in forum discussion topics differed across sub-populations: patterns for those posters with high duration of interest resemble the entire forum dynamics but were in contrast with the patterns of communication for those with high forum presence. We unpack each of the findings from Table 3 to explain these findings.

Social-Task and Social Non-Task Discussions. Forum presence was positively related to posting to both social task and social non-task discussions. In particular, forum presence has a statistically significant log coefficient of 0.29. This log coefficient is interpreted as follows. For a one-week increase in the forum presence, the rate of posting in social-task discussions increased by a factor of 1.34,

- | |
|--|
| 1) (CT,4) |
| 2) (SN,1)-(CN,1) |
| 3) (CT,2)-(CN,1)-(CT,6)-(INF,1) |
| 4) (CN,1)-(INF,1)-(CN,2)-(INF,1)-(CN,2)-(CT,6) |
| 5) (CN,1)-(ST,1)-(CT,2) |

Fig. 1. Examples of sequences of posts within discussion topics (CT, CN, ST, SN, INF); each sequence encodes the chronological order of discussion topics one student contributed to within a course discussion forum. The example shows that the first learner made 4 posts during the entire course, and their sequence of posts (reflective of discussions she engaged in) was CT-CT-CT-CT. Learner in row five made a total of four posts, but her sequence was CN-ST-CT-CT.

Table 2Number of sequences (N_{seq}) and posts (N_{posts}) in each of the two outlier groups and the group subjected to clustering.

	One-timers	Super-residents		Regular cases - used for clustering	
	$N_{\text{seq.}} = N_{\text{posts}}$	N_{seq}	N_{posts}	N_{seq}	N_{posts}
Engineering 1	671	17	2073	599	3031
Programming	366	15	2830	482	3044
Engineering 2	514	15	1887	468	2420
Data	728	18	1266	655	3929

Table 3

Outputs of negative binomial regression models.

Variables	Social Task			Social Non-Task		
	B (SE)	IRR	CI	B (SE)	IRR	CI
Constant	***-3.79 (0.14)	0.02	0.02–0.03	***-1.09 (0.12)	0.34	0.26–0.43
Forum presence	*** 0.29 (0.07)	1.34	1.16–1.54	*** 0.23 (0.03)	1.26	1.19–1.34
Middle of the course	−0.07 (0.19)	0.93	0.64–1.35	***-2.08 (0.11)	0.12	0.10–0.16
End of the course	0.01 (0.24)	1.01	0.62–1.63	***-1.70 (0.15)	0.18	0.14–0.25
Duration of interest	−0.01 (0.06)	0.99	0.89–1.11	**−0.06(0.02)	0.94	0.90–0.98
High teacher presence	*** 0.86 (0.09)	2.37	1.98–2.82	**−0.39 (0.13)	0.68	0.52–0.88
High community	0.07 (0.09)	1.08	0.90–1.29	*** 0.61 (0.13)	1.85	1.41–2.42
Forum presence*Middle	*** 0.44 (0.10)	1.55	1.27–1.90	*** 0.46 (0.06)	1.59	1.42–1.80
Forum presence*End	. 0.14 (0.08)	1.16	0.97–1.38	*** 0.31 (0.05)	1.36	1.22–1.51
Duration of interest*Middle	**−2.6 (0.08)	0.77	0.65–0.91	***−0.23 (0.05)	0.80	0.72–0.88
Duration of interest*End	−0.01 (0.07)	0.98	0.85–1.14	**−0.15 (0.04)	0.86	0.78–0.94
Alpha (over-dispersion)	0.3			1.17		
Measures of Fit	N = 7269 LL: 2456.8 AIC: 4939 BIC: 5029			N = 7269 LL: 5292 AIC:10611 BIC:10701		

	Content Task			Content Non-Task			Informational		
	B (SE)	IRR	CI	B (SE)	IRR	CI	B (SE)	IRR	CI
Constant	***-2.59 (0.17)	0.07	0.05–0.11	***-1.29 (0.31)	0.27	0.15–0.50	***-2.19 (0.16)	0.11	0.08–0.15
Forum presence	*** 0.39 (0.03)	1.49	1.39–1.59	*** 0.36 (0.03)	1.43	1.34–1.52	*** 0.45 (0.04)	1.58	1.44–1.73
Middle of the course	*** 0.89 (0.08)	2.45	2.09–2.88	*-0.17 (0.08)	0.84	0.71–0.99	−0.09 (0.11)	0.91	0.72–1.15
End of the course	*** 1.1 (0.09)	3.01	2.48–3.66	−0.02 (0.11)	0.97	0.78–1.21	0.22 (0.14)	1.25	0.93–1.68
Duration of interest	-.05 (0.02)	0.95	0.90–1.01	*-0.05 (0.02)	0.95	0.90–1.00	***-0.12 (0.03)	0.88	0.82–0.94
High teacher presence	*** 0.95 (0.19)	2.61	1.79–3.79	0.07 (0.35)	1.07	0.53–2.15	0.04 (0.17)	1.05	0.74–1.48
High community	* 0.44 (0.19)	1.56	1.07–2.27	0.18 (0.35)	1.21	0.60–2.42	*0.4 (0.17)	1.50	1.06–2.11
Forum presence*Middle	** 0.11 (0.04)	1.12	1.03–1.22	*** 0.26 (0.04)	1.30	1.19–1.43	.0.12 (0.06)	1.14	1.00–1.30
Forum presence*End	0.06 (0.04)	1.07	0.98–1.16	** 0.12 (0.04)	1.13	1.03–1.23	0.01 (0.06)	1.02	0.90–1.16
Duration of interest*Middle	−0.04 (0.03)	0.96	0.89–1.02	***-0.13 (0.03)	0.87	0.81–0.93	*-0.11 (0.05)	0.89	0.80–0.99
Duration of interest*End	−0.08 (0.03)	0.99	0.93–1.06	*-0.08 (0.03)	0.92	0.86–0.99	−0.04 (0.05)	0.95	0.86–1.06
Alpha (Over-dispersion)	0.8			0.7			0.325		
Measures of Fit	N = 7269 LL: 8821.7 AIC: 17669.5 BIC: 17759.1			N = 7269 LL: 8354.1 AIC: 16734 BIC: 16823			N = 7269 LL: 5165.8 AIC: 10357.5 BIC: 10447.1		

B - Unstandardized coefficient; SE - standard error; IRR - incidence rate ratio, i.e. exponentiated unstandardized coefficient; CI - confidence intervals for IRR, LL - log likelihood. All models $df = 7256$. * $p < .05$ ** $p < .001$ *** $p < .001$.

which is equal to the 34% increase in post count when all other variables are held constant. Table 3 reports log coefficients and interval rate ratios. The results are further presented as the percent in the change of posting rate, calculated from the log coefficients.

Forum presence was positively associated with the rate of posting in social non-task discussions, a 26% increase in rate for each one-week increase in forum presence.

Non-significant effect of time in the course on the rate of posts in social-tasks discussions suggests that these discussions remained stable throughout the course. However, in case of social non-task posts, significant negative coefficient for the effect of time, i.e. middle and the end of the course, suggests that social non-task discussions across the entire cohort decreased by 88% in the middle of the course from that of the beginning, and decreased by 82% as compared to the beginning of the course.

Posters with a higher forum presence were engaged in more social-task and social non-task discussions towards the middle of the course, and slightly fewer discussions towards the end. An increase of one week in forum presence was associated with a 55% increase in social task posts in the middle of the course. Posting into social non-task discussions increased 59% with a weekly increase in forum presence in the middle of the course, and 36% at the end of the course (in relation to the posting in the first two weeks).

A participants' duration of interest was negatively associated with the social non-task threads; a non-significant association was observed for social-task threads. Interaction between duration of interest and time in the course was significantly correlated with the number of both social task and social non-task discussions. In the middle of the course, the expected count of social task posts decreased by a 33% for a one week increase in the participant's duration of forum interest (to remind, long duration in interest to post on the forum is not identical to higher forum presence). For social non-task discussions, significant interaction of time with the duration of

interest in forum participation demonstrates a 20% decrease by the middle of the course, a 6% increase from the middle to the end of the course – mirroring cohort-level dynamics.

Course features were significantly associated with posting in both social task and social non-task discussions. Higher teacher presence courses had a 137% increase in social task discussions as compared to the forums without teacher presence. The opposite was the case with social non-task posts where higher teacher presence in forums was associated with a 32% decrease compared to forums without high teacher presence. In forums with observed elements of community, the rate of posting in social non-task discussions increased by 85%.

Content Task, Content Non-Task, and Informational Discussions. The effect of time on posting into content non-task discussions suggests that learners engaged in fewer of these throughout the course, with 16% decrease at the cohort-level in the middle of the

Table 4

Summary of student profiles detected through clustering of post sequences and revealed through exploratory analysis of highly infrequent (one-timers) and highly frequent (super-residents) posters. Summary statistics for forum presence, forum duration, and total posts are given as: median; (25th percentile, 75th percentile).

		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Visitors: one-timers	Super-residents
Engineering 1	Students, N	246 (19.11%)	180 (13.99%)	96 (7.46%)	77 (5.98%)	NA	671 (52.14%)	17 (1.32%)
	Presence, weeks	3; (2, 4)	3; (2, 4)	5; (4, 6)	2; (1, 3)		1	7; (6, 8)
	Duration, weeks	4; (2, 6.75)	4; (2, 7)	7; (5, 8)	2; (1, 4)		1	8; (7, 8)
	Total posts, N	3; (2, 4)	3; (2, 5)	11; (8, 17)	2; (2, 3)		1	86; (38, 148)
Programming	Visitors Interpretation	Visitors focused on CT; CN and SN -prominent over time	Visitors focused on CN though CT and SN posts are also notable	Residents with strong focus on CT topics	Visitors primarily oriented on INF topics		Focus on curriculum (CN: 23%, CT: 39%) and information (INF: 26%)	CT dominate (48%), a notable presence of INF (24%) and CN (18%)
	Students, N	102 (11.28%)	105 (12.17%)	113 (13.09%)	120 (13.90%)	42 (4.87%)	366 (42.41%)	15 (1.74%)
	Presence, weeks	3; (2, 3.75)	1; (1, 2)	2; (1, 2)	1; (1, 2)	5; (4, 5)	1	8; (7, 8)
	Duration, weeks	4; (2.25, 6)	1; (1, 3)	2; (1, 5)	1; (1, 3)	6; (5, 7)	1	8; (7, 8)
Engineering 2	Total posts, N	7; (5, 9)	2; (2, 3)	3; (2, 5)	3; (2, 5)	23; (17.25, 30.5)	1	83; (65.5, 227)
	Interpretation	Visitors focused on CT; CN, ST, INF - prominent over time	Visitors initially focused on CT, then the focus shifts to CN	Visitors with diverse sequences	Visitors with dominant orientation towards CN topics	Residents focused on curriculum (CT & CN)	Interest primarily focused on curriculum related topics (CN: 24%, CT: 43%)	Strong focus on curriculum topics (CT: 40%, CN: 42%)
	Students, N	70 (5.91%)	75 (6.33%)	254 (21.45%)	190 (16.05%)	66 (5.57%)	514 (43.41%)	15 (1.27%)
	Presence, weeks	6; (4, 7)	3; (2, 4)	2; (1, 3)	3; (2, 4)	5.5; (4, 8)	1	10; (7.25, 10)
Data analysis	Duration, weeks	9; (7, 10)	5; (3, 8)	2.5; (1, 5)	3; (2, 6)	7; (5, 9)	1	10; (8.25, 10)
	Total posts, N	12; (8, 16.75)	3; (2, 4)	3; (2, 4)	3; (2, 5)	16.5; (11, 23)	1	54.5; (45.5, 78.5)
	Interpretation	Residents; strong focus on CT; towards the end, CN and ST gained prominence	Visitors; dominant orientation towards CT	Visitors, strong focus on non-curriculum-related topics (SN & CN)	Visitors w/ dominant CN; posts become diverse towards the end	Residents focused on CN; later, CT and CN are balanced	Focus on curriculum-related topics, and especially those task-related (CN: 19%, CT: 54%)	Primary interest on non-curriculum related topics (CN: 40%, SN: 39%)
	Students, N	74 (6.10%)	164 (13.51%)	59 (4.86%)	171 (14.09%)	NA	728 (59.97%)	18 (1.48%)
Engineering 2	Presence, weeks	4; (3, 5)	2; (1, 3)	2; (2, 3)	3; (2, 4)		1	8; (6.5, 8)
	Duration, weeks	6; (5, 7)	2; (1, 5)	3; (2, 6)	4; (2, 6)		1	8; (8, 8)
	Total posts, N	12; (9, 15)	3; (2, 4)	3; (2, 4.5)	3; (2, 5.5)		1	90; (47.5, 148)
	Interpretation	Residents with strong focus on CT topics	Visitors w/ dominant orientation towards CT	Visitors w/ diverse posts, prominent CN and INF	Visitors w/ very diverse post types		CT dominate (63%), but a non-negligible presence of CN (17%) and INF (12%)	Focus on curriculum-related topics (CT: 44%, CN: 37%)

course. There was no effect of time on the rate of posting information and administrative queries, suggesting that making those posts remained stable. Finally, at the cohort-level, the rate of posting into content-task discussions increased over time: 145% increase observed by mid-course, and a further 56% increase towards the end of the course.

Forum presence was positively associated with content-task, content non-task, and informational queries. Posters increased their content-task discussions by 49% for a weekly increase in forum presence when all other variables held constant. Higher forum presence was associated with a higher rate of content non-task posts, a 43% increase for each week. The rate of poster engagement with information and administrative queries had a 58% increase associated with a unit increase in forum presence.

A student's duration of interest had a negative association with the rate of content non-task posts and informational queries. A 5% decrease in content non-task posts and a 12% decrease in informational and administrative queries were associated with a week increase in the duration of interest.

Interaction between forum presence and course timing was significantly associated with posting into content-task and content non-task discussion threads. In contrast to the dynamics observed at the cohort-level, posting content non-task discussions increased towards mid-course by 30% for each week of forum presence, and then decreased towards the end of the course (though still being at a higher level than when the course began). Likewise, posting to content-task discussions increased towards the mid of the course with higher forum presence (12% increase for a week increase in forum presence).

Interaction between duration of interest and timing of the course was also significant for content non-task and informational posts. Those with longer duration of interest engaged in fewer content non-task posts after the first two weeks. Their rate of posting informational queries also decreased by 11% after the first two weeks (for each unit increase in the length of interest). Course features were significantly associated with content-task and informational queries. Forums with high teacher presence had a 161% increase in rate of content-task discussions compared to courses lacking teacher presence. Forums with emergent learning communities had a 56% increase in content-related discussions. Posters in forums with high community development had 50% increase in informational queries over forums with low community development.

4.2. RQ2: profiling residents and visitors through communication topics

The second research question sought to identify different profiles of forum posters through the sequence analysis of communication topics. Clusters identified through agglomerative hierarchical clustering of post sequences in the four examined courses are summarised in Table 4. Posters with the highest forum presence had extremely diverse sequences whereas the remainder can be classified as visitors with particular learning needs, or those leaning towards the resident end of the forum participation continuum. All identified clusters were significantly different in the number of weeks posted (forum presence), number of weeks between the first and the last week of posting (forum duration), and the number of posts made.

Table 4 provides summary statistics (median, 25th and 75th percentiles) for forum presence, forum duration, and the total number of posts for each cluster. We also offer cluster interpretation based on the temporal distribution of post topics within the cluster's sequences. The table also includes a summary of exploratory analyses of the two outlying groups: visitors-one-timers and super-residents. The percentages in the table are computed on the total number of students who posted within a course forum, i.e. not only those whose sequences were used for clustering but also those who were removed as outliers (one-timers and super-residents).

All the clusters are significantly different from the two outlying groups (one-timers and super-residents) with respect to the features of forum presence, duration, and posting frequency. Pairwise cluster comparison with respect to the forum presence, forum duration, and frequency of posting can be summarised as follows (after applying the FDR correction for multiple testing):

- **Course 1** (Engineering 1) except for the cluster pair 1–2, all other cluster pairs are significantly different in the forum presence and duration; the effect sizes range from small to medium. As for the total number of posts, significant difference, with high effect size, is present only between cluster 3 and the other three clusters.
- **Course 2** (Programming): except for the cluster pairs 2–3, 2–4, and 3–4, all other cluster pairs significantly differ, with medium effect sizes, with respect to the forum presence and duration. The total number of posts is significantly different, with small to medium, and even high effect sizes, for all cluster pairs except the 3–4 pair.
- **Course 3** (Engineering 2): except for the cluster pairs 2–4 and 1–5, all other cluster pairs significantly differ with respect to the forum presence. All cluster pairs significantly differ in terms of forum duration, and as for the total number of posts, significant difference is present between all pairs except the 2–3 pair. For all three features, effect sizes range from small to medium to high.
- **Course 4** (Data analysis): except for the cluster pairs 2–3 and 3–4, all other cluster pairs significantly differ, with small to large effect sizes, with respect to the forum duration and the overall post count. As for the forum presence, significant difference, with small to medium effect sizes, is detected for all the pairs except the 2–3 pair.

5. Discussion

5.1. Interpretation of results

Research has identified that MOOC forum participation is characterised by distinct temporal and activity-level differences (Coffrin et al., 2014; Ferguson & Clow, 2015b; Gillani & Eynon, 2014; Hecking et al., 2016; Hill, 2013; Jiang et al., 2015; Kizilcec et al., 2013; Kovanović et al., 2019). However, few studies have undertaken inter-group comparisons of forum users, particularly in relation to the differences in the types of communication content and intentions. Understanding the differences in communication content for

individuals with different forum commitment can offer insights as to how networked learning unfolds at scale. Further, identification of posters with different commitment levels and their varying interests can help create interventions that scaffold both group and individual learning within the open network.

The present study addresses this gap by applying regressions that control for the time of posting and learner activity levels, with individual-level analysis of sequences of topic-coded forum posts. Regressions capture differences in patterns at the course level, whereas sequence analysis offers insight into the differences at the level of individual learners. The findings can help to understand resident and visitor behaviour in MOOC forums.

RQ1 inquired how communication topics in MOOC forums differed at various stages of the course, and across posters with ranging forum activity. We capture two communication patterns of different dynamics between the residents and visitors (Table 5). We observe the increase of all types of communication for the residents' group of participants. Even when the rate of postings across the entire forum declined, this group engaged in more communication across all discussion topics. In contrast, visitors' communication patterns are significantly different as the rate of posting in all communication topics declines for the visitors.

A unique finding of the study is a growing pattern for all discussion topics in *residents*. It suggests that group learning processes embedded within the forum network unfold similarly to the dynamics described in socio-constructivist literature on formal online courses. We observe here that residents increase content-related threads, increase social-task threads, and increase in social non-task threads. As we explain below, such dynamics confirm that a socio-constructivist lens is applicable in supporting the development of this sub-population of learners.

We find that communication for learners with sustained commitment largely mirror previously observed dynamics in small online groups. This important observation has not yet been demonstrated in empirical MOOC research. A similar pattern was observed in a small course, reported by Swan (2003a). Swan used affective, interactive and cohesive indicators to demonstrate the evolution of the elements of social presence in a formal course. In her study, the number of affective messages containing disclosure grew over time, comparable to social-task categories observed in our MOOC dataset that is largely comprised of self-disclosing statements. Besides the increase in self-disclosure, Swan's study (2003a) also described a decrease in the so-called cohesive indicators of social presence. Swan found that the number of cohesive indicators of social presence decreased as the course progressed. This category is similar to the social non-task discussion threads in our dataset. Curriculum-related posts underpinning this group's interactions also grow over time – this may be in line with an increase in cognitive presence. If so, such observations are comparable to the dynamics reported in formal online communities of inquiry (Arbaugh et al., 2008; Garrison & Akyol, 2013). Whereas our study does not focus on validating these, our empirical findings show that the constructs such as community of inquiry, with their cognitive and social presence, can in fact be developed and examined in MOOCs, *on the condition*, that they are applied to the resident posters' sub-group. In other words, the effectiveness of community development in a MOOC will *not* be observed through the analysis of the entirety of the forum participant population, as commonly done in MOOC studies.

Research on online university courses (formal education context), suggests that interpersonal communication tends to be lower at the commencement of a course and build over time. Such socially-oriented communications increase through affective expression and self-disclosure; as social climate and group communication is established, thereby fuelling further cognitive engagement (Hara, Bonk, & Angeli, 2000; McDonald, 1998, p. 199; Oren, Mioduser, & Nachmia, 2002; Swan, 2003a, 2003b; Walther, 1996; Walther, Anderson, & Park, 1994). We demonstrate that residents have similar dynamics. Importantly, we demonstrate that such dynamics are not captured by examining the forum at an entire cohort level (see Table 5). This is an important finding as it casts doubt on much of the existing empirical evidence on learning in MOOC forums. For instance, prior studies have found that the forum activity decreases exponentially as the course progresses (Brinton et al., 2014; Rossi & Gnawali, 2014). Our work demonstrates that this is not the case for the residents' interactions. The findings further question whether the decrease in posts is equal to the decrease in forum activity. Similarly, our analysis casts doubt on existing findings around cognitive presence in MOOC forums. Previous research has shown that the overall level of cognitive presence in forums is low (Kellogg et al., 2014; Kovanovic et al., 2018). We show that by undertaking the analyses at an entire-forum-level masks the increase in curriculum-related posts. Hence, questions remain as to whether cognitive presence increases for the resident sub-population of MOOC forums.

A final note towards interpreting residents' communication patterns as evolving communities within the networks of MOOC forum interactions should be made around the magnitude of the observed patterns. Although qualitatively the patterns may resemble those previously reported in formal online literature, the scale of these socio-cognitive processes may differ quantitatively. That is, the magnitude of changes in communication, expressed in percentage in the regression analyses, may be misleading if the base posting rate was low to start out with. That is, a 200% increase would still remain comparatively low if the base count was one post. Any decision to determine if such an increase is meaningful, is contextual. Hence, we interpret the outputs with the focus on the pattern of change (Nussbaum, Elsadat, & Khago, 2008).

The second research question sought to identify profiles of learners through sequence analysis of their discussion topics within

Table 5
Synthesis of results showing patterns of change in communication across forum sub-populations.

	Course-level Change in Communication Types				
	Content Task	Content Non-task	Social Task	Social Non-Task	Informational Queries
Residents	Increase	Increase	Increase	Increase	Increase
Visitors	Decrease	Decrease	Decrease	Decrease	Decrease
Both sub-populations	Increase	Decrease	Stable	Decrease	Stable

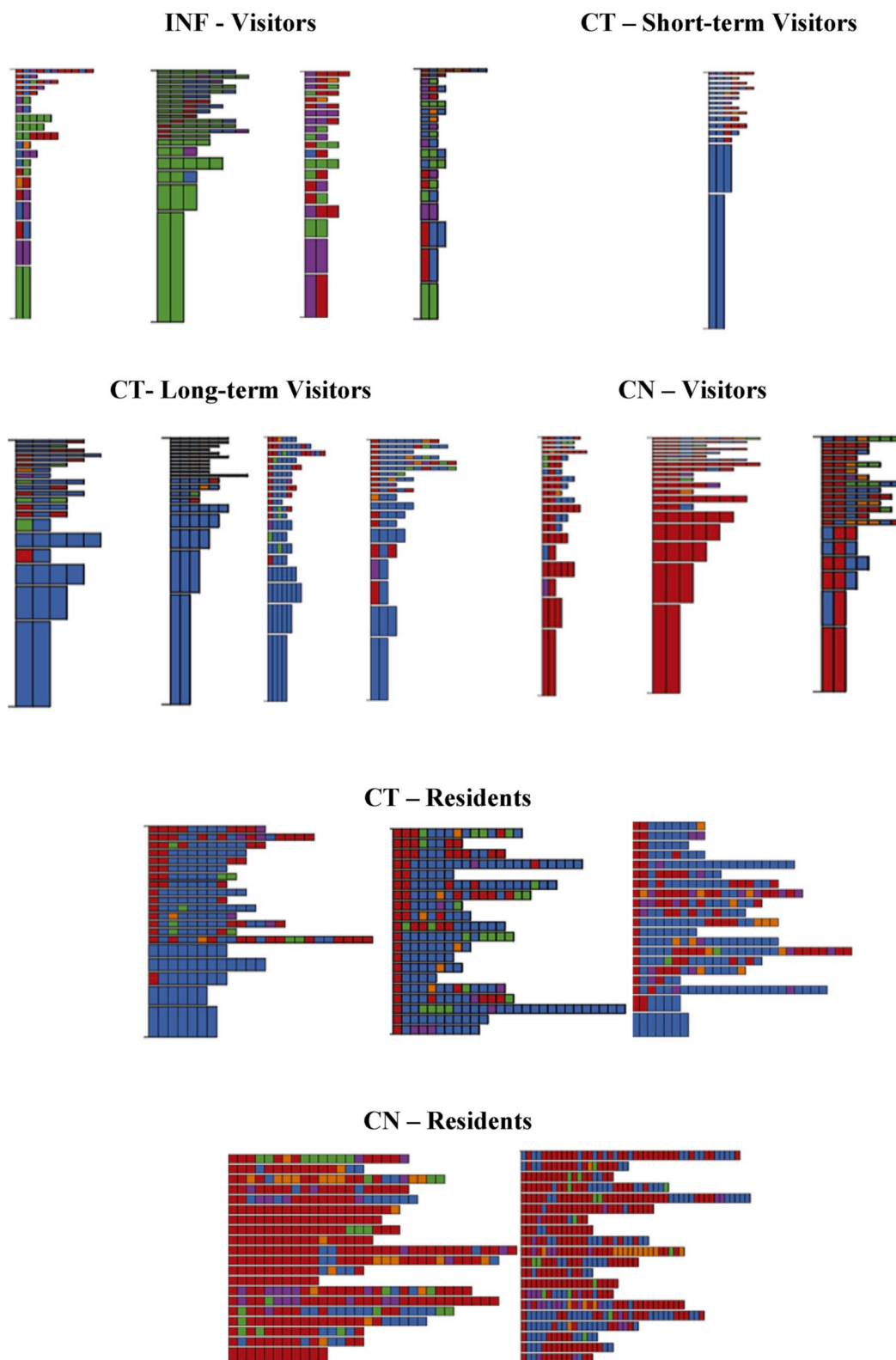


Fig. 2. Exemplary Visualisation of Sequences for Residents and Visitors. Each cluster (e.g. INF – Visitors; CT – Short-term Visitors, etc.) contains a visualisation of the cluster per course. CT contains only one visualisation since the visualisations looked similar across all four courses. Each row represents a learner, number of cells in a row represents the number of posts per person; the colour of each post represents the type of discussion (red

– CN, blue – CT, green – INF, violet – SN, orange – ST). When many learners in the cluster had the same number of sequences, and they were of the same type, the sequence of several rows is merged into one elongated cell, e.g. CT Visitors have the most prominent illustration of this. Visualisation shows the predominant type of communication in each cluster, with Residents being more diverse in their communication types, in line with the patterns observed at the course level.

MOOC forums. Fig. 2 visualizes selected sequence profiles of residents and visitors. As presented in Section 4.2, learner profiles differ in the number of posts, number of weeks posted, duration of interest to post, as well as the kinds of discourse they engage in. In line with the previous research (Wise & Cui, 2018), the main focus for those with a high commitment to forums is on the curriculum. Content-task and/or content non-task communication comprises around 80% of super-residents discourse. It is also a major focus for the residents more focused on assessment (Residents CT) or conceptual understanding and/or subject interest (Residents CN). The focus on content-task discourse is also relevant for one-timers and some visitors. Around 50% of one-timers turn to forums to engage with a content-task discussion signifying their potential struggle with assessment.

A more granular look into the composition of sequences suggests that though both residents and visitors are curriculum-driven, although they present different communication needs. We can theorize that there are two different learning orientations that overlap across these different participants: i) to pursue an individual learning trajectory and remain on the periphery or ii) to contribute to group learning activity and move to the core of the emergent group. We show that the content underlying residents and visitors with different commitments varies greatly. That is, some visitors are oriented towards assessment, whereas others are interest driven. Similarly, some residents are more concerned about assessment, whereas others are equally concerned about social-task and social non-task discussions. The presence of learners who contribute and drive group learning, as well as the presence of those who are stopping by to solve an urgent personal query suggests that personalisation of feedback and forum dashboards need to equally consider these rather distinct needs. It is also reasonable that the interventions suitable for supporting individual learning through personalisation may not be even suitable for community development within the forum.

5.2. Limitations

While the study addresses several gaps in the research literature there are some noted limitations in our work. The content qualitatively coded is not randomly sampled. That is, posters with higher presence who exclusively communicated with those with lower presence are not captured, nor are the posters with low presence who communicated exclusively with posters of low presence. Future work will address this limitation as current content labelling allows for partial automation of the labelling task. Further, the data originates from the first iteration of the four selected MOOCs. Studies have demonstrated that repeated offerings of a course can result in differing levels of forum use and topics (Poquet et al., 2018). In our view, the dynamics described in this paper are generalisable to the instances of highly desirable MOOCs that have not been offered multiple times.

Despite the study limitations, the current methodology affords an opportunity to compare the dynamics of resident posters to that observed in the vast literature on online learning. The analysis differentiates between residents and visitors beyond the engagement level (level/intensity and duration), as well as the level of content topics. Despite the limitations, the study forms an important contribution and a stepping stone towards further insights around forum sub-population analyses.

5.3. Unique contribution

The results suggest that MOOC posters with higher commitment to forums are not exclusive to the top 1% of posters with high posting levels. Characteristics associated with residents' behaviour, i.e. tendency to *be* in the social space, not only visit to solve a specific need and leave, are present among those with notably lower number of posts, such as 11–20 with the posting in 5 weeks of an eight-week course (Cluster 3, in Engineering 1 course, see Table 4).

We show that though the major focus of the residents' communication is on the curriculum, the remainder of communications for residents and super-residents is dedicated to less visible topics. Social-non task topics can be as prominent as the ones related to the curriculum (see Table 4, super-residents in Engineering 2). For residents whose bulk of communication was on curriculum posts, we found that the average number of non-content posts per person lies in the range of 3.10–4.28. Combined with the regression models, this suggests that longer commitment to forum posting is associated with diverse communication topics. This explains the findings by Wise and Cui (2018) who demonstrated that learners who posted in both content and non-content discussions had overall higher course performance.

In practical terms, residents will have less topically focused discourse, as well as the presence of core course concepts and assessment-related words. Their discourse is also likely to contain affective markers. Coupled with their sustained forum participation behaviour (and number of posts as shown in Table 4, such offers a solid ground to automated identification of this group using discourse and log data analyses.

Sequence analysis highlights the diversity of visitors to the forums. These differences are mostly observed through communication topics posters engage with, and not behavioural indicators such as post number or week counts (Table 4). Those visitors whose duration appears longer seem to diversify in their communication topics further in their sequences (e.g. cluster 1 and 2 in Engineering1 and Programming). We interpret the diversification of discourse as leaning towards the residents' side of the continuum, defined by broader spectrum of discourse topics.

Other findings relate to the association between the course features and communication patterns. We are mindful in generalising

from these as course features were compared across 4 units of analysis only. Despite the different ways teaching presence was manifested in two MOOC forums analysed, higher extent of it correlated with higher engagement in content-task discussions – consistent with evidence on the effect of teaching presence in online education literature. Further, the level of community development was positively associated with the number of content-task and social non-task communication, as well as informational queries, which is in line with theorisations around the development of learning communities in online learning literature.

6. Implications

6.1. Implications for practice

Classifying residents and visitors through the discourse and engagement features, as well as describing the patterns of communication across different sub-groups, have implications for the design and facilitation of MOOC forums. Among the most important implications for MOOC design relates to the instructional choices included to promote social learning will differ depending on the sub-population of interest. Although this statement is trivial, analysis of MOOC forums to date has not been undertaken in consideration of the differences in commitment of forum participants.

More specifically, automated technologies could support visitors with informational queries, linking them with answers from previous iteration of the same course, or chatbots offering answers to common informational concerns. That could free instructor's time to address the more cognitively complex queries by the residents or interest-driven visitors, given that bot-teachers are found ineffective to provide teaching presence such as direct instruction (Bozkurt, Kilgore, & Crosslin, 2018). Visitors driven by conceptual learning and interest (CN) would benefit from resource recommender systems that point to parts of the course, in-video segments or external resources that can satisfy their need for learning and explanations. Visitors driven by assessment and looking for content-task related are not quite interested in social learning, hence, their confusion could in part be resolved via self-explanation activities and hints to help them identify areas to dig deeper. Alternatively, matching dialogue systems can help connect visitors driven by content-task-topics with residents driven by the same interests.

Facilitation of rich experiences for forum residents should follow existing practices in online learning, requiring elements of community building and instructor or TA involvement. In other words, MOOC forum support needs to differentiate learners with sustained forum commitment, as they would benefit from direct instructor interventions. Here, automated identification of socio-emotional threads can help direct TA attention to issues that need resolution at the community level. Further, automated labelling of one's duration in forum presence in instructor-facing dashboard can help flag questions that if ignored may result in the loss of a community member. Our study links the commitment of learners with discourse features, helping design such automated systems to support learning of different sub-populations.

We show that aggregating engagement indicators across all sub-groups is likely to muddle the indicators of social learning for different sub-groups. Therefore, for evaluation of social learning, instructors need to clearly define learner sub-groups they place more priority on in a given iteration. Distinguishing between the different sub-groups of forum posters in post-course evaluations can help instructors understand the category of learners that most benefitted from instructional design, and whether that is the category they designed the course/activities for.

6.2. Implications for research

This study extends earlier work examining MOOC discussion activity. First, communication patterns across sub-groups of forum posters differ markedly from those observed at the aggregated level. We suggest that the evaluation of the group-level forum interventions from the perspective of socio-constructivist models needs to take place in relation to the forum residents and super-residents, not the entire cohort.

Second, short-term visitors have focused needs, often reflected through one communication topic. Communication of long-term visitors and residents diversifies throughout the time. Although the bulk of it remains curriculum-focused, the variance in communication topics suggests a breadth of discourse associated with a longer-term commitment. We suggest that personalisation of interventions for MOOC forums, should cater to different needs of visitor sub-groups, and account for the diversifying needs of the resident sub-groups. In other words, personalisation and forum interventions should not only to support individual learning trajectories but also group learning activities.

Third, our content analysis suggests that discourse-level features for the residents, as gleaned from natural language processing, are expected to be less cohesive due to the breadth of discourse. That is, text analysis of MOOC forums needs to account for the diversity of communication topics in the residents' sub-populations when evaluating the quality of their posts.

Fourth, the study does not capture the relationship between lurking behaviour and the behaviours examined. However, lurkers may be shifting towards visiting behaviour, or in contrast, moving towards lurking behaviour from a more explicit mode of engagement. Future work could investigate if such shifts occur, and the reasons behind them.

Finally, although our analysis is restricted to four MOOCs, and those taking place at the earlier stages of open online education, the findings highlight the need to triangulate indicators of quantity in social activity with those of quality. It is evident that interpreting the bulk of scaled interactions is not always feasible, but accounting for discourse differences, alongside time-sensitive differences in posting, is needed for a deeper understanding of communication in non-formal online education.

7. Conclusions

The future vision of open online learning is premised on networked participation. Learners with differing intentions and goals coming together in their diversity to aid learning outcomes. However, teaching practices in Massive Open Online Courses are continuously framed from the perspective of formal online education models. To address a broader need to understand MOOC forums and develop practices supporting open networked learning at scale, this study investigated forum communication dynamics between individuals with varying commitments to forum use.

The findings uncover two distinct patterns in forum communication dynamics. All types of communication decreased for forum visitors, i.e. posters mostly driven by curriculum needs, using focused either assessment-related or interest-driven discourse, and engaging with the forums for short time periods. In contrast, both socio-cognitive and informational queries increase for forum residents, i.e. posters who engaged in more diverse discourse, and though driven by curriculum-related discussions, equally engaged in social-group and personal disclosure types of communications. These so-called residents also have longer duration of forum presence and *are not* exclusively hyperposters.

The patterns of communication dynamics of residents can be qualitatively compared to those previously observed in formal online socio-constructivist studies. Importantly, our results show that the divergent dynamics between these poster sub-populations is not observed when the entire forum population is examined in its entirety, casting doubt on the claims about the lack of communities and group learning in MOOC forums.

Such offers a foundation towards scaling social teaching practices for both personalised learning as well as community development in massive open online courses. Facilitation of residents in MOOC forums should be aligned with conventional formal education practices: socioemotional processes of group formation are antecedent to further growth of posting and deeper engagement with the curriculum within the sub-group. The volume of all communication topics for visitors decreases after the initial weeks and may be better supported through more automated systems. Patterns of sub-groups are indistinguishable if their activity data is aggregated, hence planning, design, analysis and evaluation of forum activity needs to account for learner sub-groups, rather than taking a homogeneous approach.

Declaration of competing interest

None.

CRediT authorship contribution statement

Oleksandra Poquet: Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing - original draft, Visualization. **Jelena Jovanovic:** Conceptualization, Methodology, Validation, Formal analysis, Writing - original draft, Visualization. **Shane Dawson:** Conceptualization, Methodology, Validation, Writing - original draft, Supervision.

Acknowledgement

The authors thank Open Online Learning at Delft University of Technology for providing access to the data. Qualitative content analysis this study builds on, was funded by their Open Fellow Initiative in 2016.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compedu.2020.103937>.

References

- Agresti, A., & Kateri, M. (2011). *Categorical data analysis*. Springer.
- Akyol, Z., & Garrison, D. R. (2008). The development of a community of inquiry over time in an online course: Understanding the progression and integration of social, cognitive and teaching presence. *Journal of Asynchronous Learning Networks*, 12(3–4), 3–22.
- Almatrafi, O., & Johri, A. (2018). *Systematic review of discussion forums in massive open online courses (MOOCs)*. IEEE Transactions on Learning Technologies.
- Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovic, J. (2014). Engaging with massive online courses. In *WWW '14: Proceedings of the 23rd international conference on World wide web* (pp. 687–698). Scopus. <https://dl.acm.org/doi/proceedings/10.1145/2566486>.
- Arbaugh, J. B., Cleveland-Innes, M., Diaz, S. R., Garrison, D. R., Ice, P., Richardson, J. C., et al. (2008). Developing a community of inquiry instrument: Testing a measure of the Community of Inquiry framework using a multi-institutional sample. *The Internet and Higher Education*, 11(3–4), 133–136. <https://doi.org/10.1016/j.iheduc.2008.06.003>.
- Arguello, J., & Shaffer, K. (2015). Predicting speech acts in MOOC forum posts. In *Proceedings of the ninth international AAAI conference on web and social media* (pp. 2–11). University of Oxford, Oxford, UK, May 26–29, 2015. Palo Alto, California: AAAI Press.
- Atapattu, T., & Falkner, K. (2016). A framework for topic generation and labelling from MOOC discussions. In *Proceedings of the third (2016) ACM conference on learning@scale (L@S'16)* (pp. 201–204). <https://doi.org/10.1145/2876034.2893414>.
- Baker, R., Evans, B., Greenberg, E., & Dee, T. (2014). Understanding persistence in MOOCs (massive open online courses): Descriptive & experimental evidence. In U. Cress, & C. Delgado-Kloos (Eds.), *Proceedings of the European MOOC Stakeholder Summit (eMOOCs'14)* (pp. 5–10). Retrieved from <https://www.emooocs2014.eu/>.

- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2), 122–147. <https://doi.org/10.1037/0003-066X.37.2.122>.
- Boroujeni, M. S., & Dillenbourg, P. (2018). March). Discovery and temporal analysis of latent study patterns in MOOC interaction sequences. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 206–215). ACM.
- Bozkurt, A., Kilgore, W., & Crosslin, M. (2018). Bot-teachers in hybrid massive open online courses (MOOCs): A post-humanist experience. *Australasian Journal of Educational Technology*, 34(3).
- Bozkurt, A., Koutropoulos, A., Singh, L., & Honeychurch, S. (2020). On lurking: Multiple perspectives on lurking within an educational community. *The Internet and Higher Education*, 44, 100709. <https://doi.org/10.1016/j.iheduc.2019.100709>.
- Brinton, C. G., Chiang, M., Jain, S., Lam, H., Liu, Z., & Wong, F. M. F. (2014). Learning about social learning in MOOCs: From statistical analysis to generative model. *IEEE Transactions on Learning Technologies*, 346–359. <https://doi.org/10.1109/TLT.2014.2337900>.
- Chandrasekaran, M., Ragupathi, K., Kan, M.-Y., & Tan, B. (2015). Towards feasible instructor intervention in MOOC discussion forums. In *Proceedings of the international conference on information systems (ICIS'15)*. <https://aisel.aisnet.org/icis2015/proceedings/ISedu/>.
- Cheung, E. (2014). *Analyzing student engagement and retention in Georgetown's first MOOC: Globalization's winners and losers: Challenges for developed and developing Countries* (Master's thesis. Washington, DC: Georgetown University. Retrieved from <http://hdl.handle.net/10822/709735>.
- Coffrin, C., Corrin, L., de Barba, P., & Kennedy, G. (2014). Visualizing patterns of student engagement and performance in MOOCs. In *Proceedings of the fourth international conference on learning analytics and knowledge (LAK'14)* (pp. 83–92). <https://doi.org/10.1145/2567574.2567586>.
- Cramer, A. O. J., van Ravenzwaaij, D., Matzke, D., Steingrover, H., Wetzels, R., Grasman, R. P. P., et al. (2015). Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychonomic Bulletin & Review*, 1. <https://doi.org/10.3758/s13423-015-0913-5>.
- Deng, R., Benckendorff, P., & Gannaway, D. (2019). Progress and new directions for teaching and learning in MOOCs. *Computers & Education*, 129, 48–60.
- Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Lillington, NC: Edwards Brothers.
- Eynon, R., Gillani, N., Hjorth, I., & Yasserli, T. (2014). *Conceptualising interaction and learning in MOOCs. MOOC research initiative. Final report*. Retrieved in September, 2014, from http://www.moocresearch.com/wp-content/uploads/2014/06/C9146_EYNON_MOOC-Research-InitiativeEynon9146Final.pdf.
- Ezen-Can, A., Boyer, K. E., Kellogg, S., & Booth, S. (2015). Unsupervised modeling for understanding MOOC discussion forums: A learning analytics approach. In *Proceedings of the fifth international conference on learning analytics and knowledge* (pp. 146–150). <https://doi.org/10.1145/2723576.2723589>.
- Ferguson, R., & Clow, D. (2015b). Examining engagement: Analysing learner subpopulations in massive open online courses (MOOCs). In *Proceedings of the fifth international conference on learning analytics and knowledge* (pp. 51–58). <https://doi.org/10.1145/2723576.2723606>.
- Ferguson, R., & Clow, D. (2015a). Consistent commitment: Patterns of engagement across time in massive open online courses (MOOCs). *Journal of Learning Analytics*, 2(3), 55–80.
- Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, 40(4), 1–37. <https://doi.org/10.18637/jss.v040.i04>.
- Gillani, N., & Eynon, R. (2014). Communication patterns in massively open online courses. *The Internet and Higher Education*, 23, 18–26. <https://doi.org/10.1016/j.iheduc.2014.05.004>.
- Hara, N., Bonk, C. J., & Angeli, C. (2000). Content analysis of online discussion in an applied educational psychology course. *Instructional Science*, 28(2), 115–152. <https://doi.org/10.1023/A:1003764722829>.
- Haythornthwaite, C. (2019). Learning, connectivity and networks. *Information and Learning Sciences*, 120(1/2), 19–38.
- Haythornthwaite, C., & de Laat, M. (2012). Social network informed design for learning with educational technology. In A. D. Olofsson, J. O. Lindberg, K. Klinger, & C. Shearer (Eds.), *Informed design of educational technologies in higher education: Enhanced learning and teaching* (pp. 352–374). IGI Global.
- Hecking, T., Chounta, I.-A., & Hoppe, H. U. (2016). Investigating social and semantic user roles in MOOC discussion forums. In *Proceedings of the sixth international conference on learning analytics & knowledge (LAK'16)* (pp. 198–207). <https://doi.org/10.1145/2883851.2883924>.
- Hecking, T., Chounta, I. A., & Hoppe, H. U. (2017). Role modelling in MOOC discussion forums. *Journal of Learning Analytics*, 4(1), 85–116.
- Henri, F. (1992). Computer conferencing and content analysis. In A. R. Kaye (Ed.), *Collaborative learning through computer conferencing (NATO ASI series, series F: Computer and systems sciences (Vol. 90)*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-77684-7_8.
- Hill, P. (2013, March 2). *The four student archetypes emerging in MOOCs* [Web log post]. Retrieved from <http://mfeldstein.com/the-four-student-archetypes-emerging-in-moocs/>.
- Honeychurch, S., Bozkurt, A., Singh, L., & Koutropoulos, A. (2017). Learners on the periphery: Lurkers as invisible learners. *European Journal of Open, Distance and E-Learning*, 20(1), 192–212. <https://doi.org/10.1515/eurodl-2017-0012>.
- Jackman, S., Tahk, A., Zeileis, A., Maimone, C., Fearon, J., Meers, Z., & Imports, M. A. S. S. (2017). *Package 'pscl'*. Retrieved from <http://github.com/atahk/pscl>.
- Jiang, Z., Zhang, Y., Liu, C., & Li, X. (2015). Influence analysis by heterogeneous network in MOOC forums: What can we discover?. In *Paper presented at the international conference on educational data mining (EDM) (8th, Madrid, Spain, jun 26-29, 2015)*. Retrieved from <https://eric.ed.gov/?id=ED560574>.
- Joksimovic, S. (2017). *Analytics-based approach to the study of learning networks in digital education settings*. Doctoral dissertation. University of Edinburgh.
- Jones, C. (2004). Networks and learning: Communities, practices and the metaphor of networks. *Research in Learning Technology*, 12(1), 81–93. Retrieved from <http://journals.co-action.net/index.php/rlt/article/view/11227>.
- Jones, C., & Esnault, L. (2004). The metaphor of networks in learning: Communities, collaboration and practice. In *Paper presented at the 4th international conference on networked learning 2004*. UK: Lancaster University. Retrieved from <https://telearn.archives-ouvertes.fr/hal-00190280/>.
- Jones, C. (2015). *Networked learning: An educational paradigm for the age of digital networks*. Springer.
- Jones, C., & Esnault, L. (2004). The metaphor of networks in learning: Communities, collaboration and practice. *Proceedings of Networked Learning Conference*. <https://telearn.archives-ouvertes.fr/hal-00190280/>.
- Kellogg, S., Booth, S., & Oliver, K. (2014). A social network perspective on peer supported learning in MOOCs for educators. *International Review of Research in Open and Distance Learning*, 15(5). <https://doi.org/10.19173/irrodl.v15i5.1852>.
- Khalil, M., & Ebner, M. (2017). Clustering patterns of engagement in massive open online courses (MOOCs): The use of learning analytics to reveal student categories. *Journal of Computing in Higher Education*, 29(1), 114–132.
- Kizilcec, R., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of the third international conference on learning analytics and knowledge (LAK'13)* (pp. 170–179). <https://doi.org/10.1145/2460296.2460330>.
- Kizilcec, R., & Schneider, E. (2015). Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale. *ACM Transactions on Computer-Human Interaction (TOCHI) - Special Issue on Online Learning at Scale*, 22(2), 1–24. <https://doi.org/10.1145/2699735>. Art.6.
- Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., & Adesope, O. (2015). Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. *The Internet and Higher Education*, 27, 74–89. <https://doi.org/10.1016/j.iheduc.2015.06.002>.
- Kovanovic, V., Joksimovic, S., Poquet, O., Hennis, T., Cukic, I., de Vries, P., & Gasevic, D. (2018). Exploring communities of inquiry in massive open online courses. *Computers & Education*, 44–58. <https://doi.org/10.1016/j.compedu.2017.11.010>.
- Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., de Vries, P., Hatala, M., et al. (2019). Examining communities of inquiry in Massive Open Online Courses: The role of study strategies. *The Internet and Higher Education*, 40, 20–43.
- Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., et al. (2016). Towards automated content analysis of discussion transcripts: A cognitive presence case. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 15–24). ACM.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Sage publications.
- La Vista, D., Falkner, N., & Szabo, C. (2017). Understanding the effects of intervention on computer science student behaviour in on-line forums. In *Proceedings of the 2017 ACM conference on innovation and technology in computer science education* (pp. 200–205). ACM.
- Littlejohn, A., & Hood, N. (2018). *Reconceptualising learning in the digital age: The [un] democratising potential of MOOCs*. Singapore: Springer.
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables* (Vol. 7). Thousand Oaks: Sage Publications.
- Luna, J. M., Fardoun, H. M., Padillo, F., Romero, C., & Ventura, S. (2019). Subgroup discovery in MOOCs: A big data application for describing different types of learners. *Interactive Learning Environments*, 1–19. <https://doi.org/10.1080/10494820.2019.1643742>.

- Magnusson, A., Skaug, H., Nielsen, A., Berg, C., Kristensen, K., Maechler, M., et al. (2017). Package "glmmTMB." R package version 0.2.0. <https://cran.microsoft.com/web/packages/glmmTMB/glmmTMB.pdf>.
- McDonald, J. (1998). *Interpersonal aspects of group dynamics and development in computer conferencing*. Doctoral dissertation. The University of Wisconsin – Madison (Retrieved from Proquest Database).
- Moore, R. L., Oliver, K. M., & Wang, C. (2019). Setting the pace: Examining cognitive processing in MOOC discussion forums with automatic text analysis. *Interactive Learning Environments*, 1–15.
- Nussbaum, E. M., Elsadat, S., & Khago, A. H. (2008). Best practices in analyzing count data: Poisson regression. *Best Practices in Quantitative Methods*, 306–323.
- Oren, A., Mioduser, D., & Nachmia, R. (2002). The development of social climate in virtual learning discussion groups. *International Review of Research in Open and Distance Learning*, 3(1). <https://doi.org/10.19173/irrodl.v3i1.80>.
- Poquet, O. (2017). Doctoral. *Social context in MOOCs*. University of South Australia.
- Poquet, O., & Dawson, S. (2016). Untangling MOOC learner networks. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 208–212. <http://dl.acm.org/citation.cfm?id=2883919>.
- Poquet, O., Kovanović, V., de Vries, P., Hennis, T., Joksimović, S., Gašević, D., & Dawson, S. (2018). Social presence in massive open online courses. *International Review of Research in Open and Distributed Learning*, 19(3).
- Ramesh, A., & Getoor, L. (2018). Topic evolution models for long-running MOOCs. In *International conference on web information systems engineering* (pp. 410–421). Springer.
- Rossi, L. A., & Gnowali, O. (2014). Language independent analysis and classification of discussion threads in Coursera MOOC forums. In *Information reuse and integration (IRI), 2014 IEEE 15th international conference on* (pp. 654–661). <https://doi.org/10.1109/IRI.2014.7051952>.
- Seaton, D. T., Bergner, Y., Chuang, I., Mitros, P., & Pritchard, D. E. (2014). Who does what in a massive open online course? *ACM Communications*, 57. <https://doi.org/10.1145/2500876>.
- Shirvani Boroujeni, M., Hecking, T., Hoppe, H. U., & Dillenbourg, P. (2017). Dynamics of MOOC discussion forums. In *Proceedings of the 7th international learning analytics and knowledge conference (LAK17)*, Vancouver, Canada. Retrieved from <https://infoscience.epfl.ch/record/223718>.
- Siemens, G. (2013). Massive open online courses: Innovation in education? In R. McGreal, W. Kinuthia, & S. Marshall (Eds.), *Open educational resources: Innovation, research and practice* (pp. 5–15). Athabasca: Commonwealth of Learning, Athabasca University.
- Studer, M., & Ritschard, G. (2016). What matters in differences between life trajectories: A comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society: Series A*, 179(2), 481–511. <https://doi.org/10.1111/rssa.12125>.
- Stump, G. S., DeBoer, J., Whittinghill, J., & Breslow, L. (2013). Development of a framework to classify MOOC discussion forum posts: Methodology and challenges. In *[TLL working paper]*. Retrieved from https://tll.mit.edu/sites/default/files/library/Coding_a_MOOC_Discussion_Forum.pdf.
- Sun, N., Rau, P. P.-L., & Ma, L. (2014). Understanding lurkers in online communities: A literature review. *Computers in Human Behavior*, 38, 110–117. <https://doi.org/10.1016/j.chb.2014.05.022>.
- Swan, K. (2003a). Developing social presence in online course discussions. In S. Naidu (Ed.), *Learning and teaching with technology: Principles and practices* (pp. 147–164). UK: Routledge.
- Swan, K. (2003b). Learning effectiveness: What the research tells us. In J. Bourne, & J. C. Moore (Eds.), *Elements of quality online education, practice and direction* (pp. 13–45). Needham, MA: Sloan Center for Online Education.
- Tang, H., Xing, W., & Pei, B. (2018). Exploring the temporal dimension of forum participation in MOOCs. *Distance Education*, 39(3), 353–372. <https://doi.org/10.1080/01587919.2018.1476841>.
- Walther, J. (1996). Computer-mediated communication impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, 23(1), 3–43. <https://doi.org/10.1177/009365096023001001>.
- Walther, J., Anderson, J. F., & Park, D. W. (1994). Interpersonal effects in computer-mediated interaction: A meta-analysis of social and antisocial communication. *Communication Research*, 21(4), 460–487. <https://doi.org/10.1177/009365094021004002>.
- Wang, X., Wen, M., & Rosé, C. P. (2016). Towards triggering higher-order thinking behaviors in MOOCs. In *Proceedings of the sixth international conference on learning analytics & knowledge (LAK'16)* (pp. 398–407). <https://doi.org/10.1145/2883851.2883964>.
- Wang, X., Yang, D., Wen, M., Koedinger, K., & Rosé, C. P. (2015). Investigating how student's cognitive behavior in MOOC discussion forums affect learning gains. In O. Santos, J. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, et al. (Eds.), *Proceedings of the 8th international conference on educational data mining (EDM'15)*, Madrid, Spain (pp.226–223). Retrieved from <http://www.educationaldatamining.org/EDM2015/proceedings/full226-233.pdf>.
- White, D. S., & Le Cornu, A. (2011). Visitors and residents: A new typology for online engagement. *First Monday*, 16(9). Retrieved from <http://firstmonday.org/ojs/index.php/fm/article/viewArticle/3171>.
- Wise, A., & Cui, Y. (2018). Learning communities in the crowd: Characteristics of content related interactions and social relationships in MOOC discussion forums. *Computers & Education*, 122, 221–242. <https://doi.org/10.1016/j.compedu.2018.03.021>.
- Wise, A., Cui, Y., Jin, W., & Vytasek, J. (2017). Mining for gold: Identifying content-related MOOC discussion threads across domains through linguistic modeling. *The Internet and Higher Education*, 32, 11–28. <https://doi.org/10.1016/j.iheduc.2016.08.001>.
- Wise, A., Cui, Y., & Vytasek, J. (2016). Bringing order to chaos in MOOC discussion forums with content-related thread identification. In *Proceedings of the sixth international conference on learning analytics & knowledge (LAK'16)* (pp. 188–197). <https://doi.org/10.1145/2883851.2883916>.
- Yang, D., Wen, M., Kumar, A., Xing, E. P., & Rosé, C. (2014). Towards an integration of text and graph clustering methods as a lens for studying social interaction in MOOCs. *International Review of Research in Open and Distance Learning*, 15(5). <https://doi.org/10.19173/irrodl.v15i5.1853>.
- Zawacki-Richter, O., Bozkurt, A., Alturki, U., & Aldraiweesh, A. (2018). What research says about MOOCs—An explorative content analysis. *International Review of Research in Open and Distance Learning*, 19(1).