Learner Communications in Massively Open Online Courses

Nabeel Gillani[1]

September 1, 2013

Abstract

Massively open online courses (MOOCs) bring together hundreds to thousands of people from around the world to learn a variety of topics in short (2 to 15) week courses. Until now, most have not offered formal institutional credit but have been freely available to anyone with an internet connection, regardless of their educational background. MOOCs have become a popular topic in higher education largely because they enable a geographically diverse group of learners to access educational resources from the world's top universities. They have evolved from previous incarnations of online learning but are distinguished in their global reach and semi-synchronicity.

In the past two years, MOOCs have received very polarized media attention. Some believe MOOCs will completely transform traditional models of higher education. Others view them as mechanisms for furthering a commoditization of learning that is best experienced in small groups and in-person. Unfortunately, a great deal of this debate has lacked theoretical grounding and evidence from rigorous research. Sound investigation is needed to move beyond these extreme views and evaluate the true pedagogical potential of MOOCs.

This work analyses a key differentiator of MOOCs from previous efforts at open education – communication between a global body of learners via online discussion forums – to discover who tends to interact online and how. Literature on MOOCs has not yet indicated the backgrounds, motivations, or achievement levels of forum participants. It also has not revealed their communication patterns or how groups tend to form and disband around certain topics. As MOOCs enable communication between learners that may have otherwise never interacted, it is essential to gain insights into how they engage in online discussions to better support their learning.

This study aims to address this need. Data analysis of nearly 87,000 individuals from a case study of a particular MOOC reveals a number of key trends. Forum participants – like those in the course more broadly – tend to be young adults from the western world. Students in the course favour "real-world" topics that have relevance and significance in their lives beyond the academic setting. Forum participants assemble and disperse quickly as crowds, not communities, of learners. Finally, those that engage explicitly in the discussion forums are often higher-performing than their counterparts in the course, although the vast majority of forum participants receive "failing" marks. These findings have implications for how certain types of MOOCs may encourage and promote online discussions in the future – and how these discussions can help students learn.

---

[1] MSc Candidate, Department of Education, University of Oxford. nabeel.gillani@new.ox.ac.uk.

Introduction

In 2011, Peter Norvig and Sebastian Thrun of Stanford University offered an introductory class on Artificial Intelligence (AI)[2]. However, the course, previously a staple of Stanford's engineering and computer science curricula, had a twist: anyone in the world could enrol. At the peak of registration, over 200,000 students from around the world were present to learn the fundamentals of machine learning, natural language processing, and other core concepts in statistics and engineering. For those enrolling outside of Stanford, there were no entry assignments, no costs incurred, and no formal credit offered.

Norvig and Thrun's AI course is credited for launching the current boom in massively open online courses, or MOOCs for short. MOOCs are examples of open courseware – freely available content from education institutions, distributed via the internet – that enable hundreds to thousands of individuals from around the world to interactively learn with one another (Littlejohn, 2013). MOOCs tend to have a defined start and end date and the majority of courses range between 2 to 15 weeks in length. They are "open" in that they often do not have a registration capacity or formal educational pre-requisites: anyone with an internet connection can join. There are two major types of MOOCs: xMOOCs and cMOOCs. The former generally have course structures that are predetermined by an instructor and tend to have regular homework sets, lectures, and exams similar to traditional "brick-and-mortar" institutions. The latter focus on how large-scale interactive learning environments can be used to connect participants to enable them to explore mutual areas of interest, often through discussion forums (Littlejohn, 2013; McAuley, et al., 2010).

In the past 2 years, MOOCs have been a subject of both hope and concern. Advocates like Thomas Friedman have enthusiastically labelled MOOCs as the revolutionary forces that will disrupt traditional practices in higher education[3] and others have begun to investigate how they may help people in the developing world gain access to education (Ryan, 2013). Others closer to the classroom, like philosophy professors at San Jose State University, have expressed public concern and frustration over the displacement of diverse, personalized pedagogy with one-size-fits-all courseware[4], reminiscent of David Noble's prophecy of online learning as a harbinger of "Digital Diploma Mills" (Nobel, 1998).

If the evolution of MOOCs follows Gartner's hype cycle (Fenn, et al., 2009), then the polarised response to such courses is to be anticipated. Media coverage of MOOCs has fuelled both hype and qualified disillusionment as thought leaders across disciplines offer their perspectives on the future of higher education. Until now, many of these discussions have not been grounded in educational theory or a systematic, rigorous investigation of the potential of MOOCs. To help move beyond this undue polarisation, this research aims to explore a key dimension of MOOCs – how a diverse body of learners engage with one another through online discussions – in order to explore the true pedagogical potential of this new courseware.

There have been early attempts to analyse and understand MOOCs. One focus of this research has been the low completion rates characteristic of many MOOCs. Recent research from Stanford has highlighted the different ways in which MOOC participants tend to engage with coursework in massively open learning settings, stressing that MOOCs should not simply be assessed or interpreted in terms of high attrition rates or students' varying degrees of formal participation (Kizilcec, et al., 2013).

---

[2] https://www.ai-class.com/overview

[3] http://www.nytimes.com/2013/01/27/opinion/sunday/friedman-revolution-hits-the-universities.html?pagewanted=2

[4] http://chronicle.com/article/The-Document-an-Open-Letter/138937/

Other researchers have started to investigate how educators may learn from MOOCs in order to improve students' "offline", campus-based experiences. Recent work from campuses affiliated with the edX MOOC platform illuminated how some MOOC participants that supplemented their learning with offline collaborations received higher marks than those that solely participated online (Breslow, et al., 2013). While the data is early-stage, these insights are seeding an ongoing investigation into if, and how, MOOCs may help augment existing practices in higher education (Breslow, et al., 2013).

These early studies have leveraged fine-grained data collected on learners' participation and mutual interactions. These "digital traces" that were once virtually impossible to capture in campus-based contexts – such as the number of students that opened a textbook before the final exam, or which students spoke to which other students about the final problem set – are now available at unprecedented levels of detail and can be mined for deeper insights into how people learn. Even so, learners from around the world have used open educational resource repositories like MIT OpenCourseWare and iTunesU for years to download and view lectures, complete online quizzes and tests, and assess their understanding of content in various other ways. These sites already have the capacity of supporting large, global user bases and logging their platform interactions. What, then, makes MOOCs different from their predecessors in online learning? Never before has such a large body of people assembled together in a classroom. If MOOCs are today's global online classrooms, then they are different precisely because they promote interactions between these learners in a semi-synchronous course setting.

 To begin to understand what MOOCs mean for education it is important to first learn how participants in these settings come together, discuss concepts, and form communities. Some students in MOOCs spend as much or more time on using the discussion forums as they do viewing lectures or doing homework (Seaton, et al., 2013). While most MOOCs so far have focused on conveying university-level content, their true potential perhaps lies in enabling the creation of new knowledge and insights through online learner interactions.

This work aims to build upon earlier research on the formation and evolution of online communities. Stahl's theories in computer-supported collaborative learning (Stahl, et al., 2006) have inspired investigations into how learning occurs via online discussions on a variety of formal and non-formal educational platforms. It is perhaps these discussions in MOOCs - conversations that are guided by course content, between thousands of people from around the world, each with unique life experiences and sociocultural contexts - that separate them from previous distance learning efforts.

To understand the nature of these conversations it is important to understand the learners that partake in them. Duke University's recent review of its bioelectricity MOOC sheds some light on the different types of people that take MOOCs. Out of over 3,500 students that took the pre-course survey, over 56% were over the age of 26, 72% already had at least a Bachelor's degree, and over 40% of people said that future job preparation was an important reason for taking the course (Belanger & Thornton, 2013). Another paper on MIT's "Circuits and Electronics" MOOC found that over 60% of survey respondents already had at least a Bachelor's degree. Moreover, over 55% took the course primarily because they wanted to gain new skills and knowledge (Deboer, et al., 2013).

As studies on MOOCs continue to emerge, they will help paint a more complete picture of who tends to take these courses and how users' characteristics differ according to the types of courses that are offered. Still, through data from the Duke and MIT MOOCs, it is readily apparent that those who enrolled do not have the profile of a "typical" undergraduate student. These participants are young but still old enough to have had one or two full-time jobs, many are lifelong learners that already possess undergraduate or advanced degrees, and some are looking at the MOOC as a way to enhance their professional credentials and skillsets. In other words, they have unique life

experiences and motivations that they carry with them to the course. It is the discussion forums in a MOOC, then, that enable them to begin to share their ideas and insights with their peers.

But to really learn about these learners, it is important to see how they behave in settings that encourage applied learning and problem solving. This follows directly from the fact that large proportions of MOOC participants already have post-secondary degrees and are now in professional training programs or in the workforce - i.e., are members of the "real-world". Here, our notion of "real-world learning opportunities" in a massively open education setting relates closely to the applied problem-solving components characteristic of many problem-based learning settings (Savery & Duffy, 1995). Learning how people engage with each other around topics that are relevant to their lives outside of the course may help educators design more effective learning environments and experiences in the future.

This study examines early data from a MOOC with an emphasis on real-world learning in order to generate initial insights into who tends to use the discussion forums, how, and for which purposes. This work is primarily exploratory in nature; it leverages observations, surveys, and social network analysis to gain insights into the structure attributes of communication between learners. It also explores the relationships between forum participation and demographics. Finally, because many learners appear to take MOOCs for different reasons and with different levels of formal engagement, it investigates how forum participate relates to the students' final marks in the course. The investigation is challenging because it requires both computational and qualitative techniques to extract insights from a large-scale dataset of fine-grained user behaviours and attributes. However, it is important because MOOC research is in its infancy and there is currently little understanding of how learners communicate with one another in these settings. Insights into these discussions could lead to new discoveries in how people gain and exchange knowledge.

The rest of the study is organized as follows. Section 2 draws upon the key educational and sociocultural theories that underlie online forum discussions and applied learning in order to present the study's primary research questions. Section 3 introduces the MOOC on business strategy that serves as a case study for this research and describes the computational and qualitative techniques used to perform analysis of its forum discussions. Section 4 presents the results of the investigation. Section 5 discusses these results in the context of previous research into online learning communities and other social networks and offers some recommendations for MOOC instructors. Section 6 concludes the piece and discusses opportunities for future research.

Supporting Literature

The evolution of distance education

Massively open online courses are one product of nearly 200 years of evolution in distance education. As Larreamendy-Joerns and colleagues describe in "Going the Distance with Online Education", distance education began with courses by mail in the 19th century where instructors would send syllabi to students and receive completed assignments through the post. The popularization of mass media such as radio and television helped extend educational opportunities to geographically dispersed populations (Larreamendy-Joerns & Leinhardt, 2006).

This desire to increase access through various media reflects a growing propensity to distribute, at a larger scale, educational content that was once confined to the brick-and-mortar institution. With the advent of public internet browsers and improvements in internet connectivity in the mid-1990's, the internet became a popular medium for disseminating educational

resources. Initiatives like the Open University in the UK, The International School of Information Management in Santa Barbara, California, and Nova University in Florida are a few examples of early efforts to use the internet to share educational content and connect learners through discussion forums that promote peer engagement (Bork, 1995).

Perhaps one of the most emphatic announcements in distance education came in 2001 when MIT launched its OpenCourseWare (OCW) initiative. Envisioned during the dot-com era - when many universities were racing to form alliances that would enable them to leverage the internet to deliver education to the masses (not unlike the current MOOC environment) - OCW launched as a free-to-use online repository of MIT's lecture videos, notes, problem sets, exam papers, and other resources available to anyone in the world with an internet connection. By 2005, the MIT-based initiative had expanded to become the OCW consortium, comprised of over 100 institutions of higher education committed to putting up content from their own classrooms (Abelson, 2008). As public awareness of OCW spread, many people began to question the role of higher education institutions if content could be freely accessed from anywhere. MIT - and other participating universities - defended the on-campus experience as beyond what could be packaged and made available for download.

Spikes in consumer demand for portable audio players in the early 2000's also encouraged the rise of additional open educational resource platforms like iTunes U, where learners could download lectures, often for free, from a variety of institutions across a number of different disciplines. In the past few years, some researchers have investigated platforms like iTunes U in order to gain insights into if, and how, podcasts could lead to more positive educational outcomes for learners in comparison to traditional professor-driven lectures delivered in-person (McKinney, et al., 2008). Other platforms like Carnegie's Open Learning Initiative have aimed to leverage research and insights from cognitive science and statistics to create more personalized online learning experiences (Johnstone, 2005). The financial sustainability of these platforms has been readily called into question, inciting discussions about how they may continue to offer their services to a global audience of learners (Wiley, 2007).

MOOCs can be viewed as the newest incarnations of these earlier attempts at distance education, in part because of the technological and financial challenges they face, but also because of the discussions they've inspired about how learners' "online" experiences compare to "offline", campus-based education. In some sense, then, the MOOC story has already been told. So, how are MOOCs different from their predecessors? MOOCs are hybrids of previous attempts at online distance education: they bring together early approaches to online learning and the scale and potential reach of open courseware efforts. Many online learning efforts started off as being synchronous, and therefore, engendered group discussions, but the scale of these discussions was small and geographically constrained. The rise of global-scale open educational resource repositories like OCW and iTunesU offered access to content "to the masses" but failed to create a space, at least at their onset, where learners could engage with new content *together*[5]. While MOOCs are not entirely synchronous in comparison to most offline settings since they enable learners to watch lectures and participate in discussions according to their own schedules, they do have a defined start and end date, and many encourage students to partake in weekly assignments as the course progresses (Kizilcec, et al., 2013). This semi-synchronicity - and the interactions it

---

[5] People could access lectures on-demand at any point in time and did not, until years after some platforms launched, have an outlet for engaging in group discussions around the content.

promotes among a diverse group of learners – is the focus of this study largely because of its importance in the context of educational and sociocultural theories of learning.

Sociocultural underpinnings of learning at a massive scale

Despite the immense amount of attention devoted to MOOCs in recent months, few articles have discussed how these new manifestations of open courseware reflect existing conceptions of how learners learn, despite the fact that they capture elements of many different sociocultural theories. One of the most relevant frameworks underlying MOOCs is social constructivism. Pioneered by Lev Vygotsky in the early 20th century, social constructivism argues that a learner's potential to learn is not realized in isolation, but rather, through conversations and interactions with other learners. These interactions can help learners explore their unique zones of proximal development, or their capacities to gain and share knowledge as their ideas are drawn out and reinforced with the help of others (Vygotsky, 1978).

There is much literature that discusses the role of social interactions in learning settings. Ravenscroft draws from the work of Bakhtin as he describes "dialogic" – the clash of opinions and meaning (e.g. semantics) – instead of "dialectic" – the formal tools used to communicate (e.g. syntax) – as fundamental to the social construction of knowledge (Ravenscroft, et al., 2007; Bakhtin, 1986). In a similar vein, some researchers have investigated the process of "identity negotiation" in peer learning settings and how the individual's role and subsequent learning is influenced when contextualized with other learners (Hughes, 2010). Others have aimed to define the long-debated notion of "group learning" (Wilson, et al., 2007). In the context of MOOCs, the notion of group learning as it occurs in online discussions is of particular interest. Indeed, since the late 1990's, researchers have used the ideas that underpin social constructivism to explore how learners gain and co-create knowledge in online discussion forums. In part, their studies have aimed to improve the design of existing web-based discussion environments based on what encourages "meaningful", or learning-inducing, discussions among participants (Lefoe, 1998; Woo & Reeves, 2007). They have also taken a granular look at exactly how communication between learners in online communities differs by defining metrics for how discussion participation varies across groups of different sizes (Nielsen, et al., 2010). In many cases, however, the datasets that have motivated these investigations have been sparse or small-scale, making it difficult to gain a holistic understanding of how people engage and learn through their online interactions, and why.

Social constructivism does not capture all of the different "types" of learning that MOOCs engender. Some scholars have analysed how current MOOCs employ elements of behaviourism in the way they leverage auto-graded assignments as learning mechanisms (Daniel, 2012). Still, we can argue that social constructivism is a guiding theoretical framework for understanding at least one type of learning - social learning - that some MOOCs promote. Even so, social constructivism must be coupled with another sociocultural theory in order to capture an important part of massive-scale online learning: the richness of experience and differences of thought that characterize its culturally diverse pool of learners. To begin to investigate discussions in MOOCs, it is not enough (or perhaps, even accurate) to simply assume that MOOC participants are integral contributors to each others' educational experiences. Instead, it is important to also understand how their unique backgrounds and motivations colour their interactions online.

Activity theory is one tool that underlies this investigation. This sociocultural theory, envisioned by the Soviet Psychologist A. N. Leont'ev, is similar to social constructivism in its emphasis on the individual's role in her larger environment, and how her interactions with this and other environments impact both individual and societal development (Leont'ev, 1978). Activity

theory suggests that the way human beings interact with the world is inherently shaped by their history, culture, and a number of other environmental factors. Since Leont'ev's pioneering work, many researchers have explored the application of activity theory to new domains. In particular, Engestrom's work on expansive learning has built upon this theory by investigating how people (particularly those in the workforce) use their past experiences to envision new, innovative ways of solving problems (Engestrom, 2008).

The ideas underlying activity theory help motivate the need to investigate online discussions in massively open educational settings. MOOC participants are culturally and geographically diverse, and each person's set of previous life experiences offers something new to a mix of thousands of others. Activity theory, then, provides an additional set of lenses through which we can understand how communication both among and between these different types of people occurs. In tandem, both social constructivism and activity theory justify the importance of considering not just how learners can help each other learn, but also, how the different backgrounds and attributes of these learners combine to shape the ways in which they communicate and engage with each other.

Crowds and communities of online learners

While social constructivism and activity theory highlight how interactions among a diverse group of learners can impact individual learning processes, they do not, at their core, account for the role of technology in helping to facilitate these interactions. Stahl's concept of computer-supported collaborative learning (CSCL) mitigates this by analysing how learners can collaboratively create knowledge as they use computers. Stahl is explicit, however, in discouraging the use of CSCL as an umbrella framework to use when analysing learners in online educational contexts, as many of these settings tend to be individual rather than group-focused (Stahl, et al., 2006). In this way, he distinguishes between cooperation in learning settings, where individuals are largely responsible for working on their own within a group setting in order to gain knowledge, and collaboration, which, as Roschelle and Teasley explain, involves group-directed negotiation and the construction of shared goals and meaning (Roschelle & Teasley, 1995).

Stahl's distinction between cooperation and collaboration in computer-supported learning has analogues in problem-solving literature. Caroline Haythornthwaite's research in crowdsourcing[6] has yielded a dichotomous classification of problem solvers as either being "lightweight" or "heavyweight" contributors. While lightweight participants tend to have weaker ties to other participants and contribute solutions that are more ad-hoc and fragmented, "heavyweights" seek a stronger community around problem-solving and tend to engage more consistently with their peers and the task at hand (Haythornthwaite, 2009). In this way, both knowledge "crowds", comprised of lightweight contributors, and "communities", more permanent structures often grounded in the participation of key heavyweights, emerge in online settings (Haythornthwaite, 2009). Members of the crowd, like cooperative learners, tend to participate individually[7] in a broader group setting; community members, on the other hand, often engage in the types of iterative and negotiated dialogue that leads to the collaborative creation of new knowledge.

---

[6] Crowdsourcing as a distributed, online problem-solving methodology has gained popularity in the past 15 years, particularly through commercial challenges like those posed by companies like Threadless, as well as research and development problems put forth by platforms like Innocentive **Invalid source specified.**.

[7] It is important to note that acting "individually" in this case is not synonymous with acting "selfishly" -- indeed, a recent paper noted that lightweight participants tend to be "intrinsically" motivated to contribute to crowdsourcing (Budhathoki & Haythornthwaite, 2012). Instead, "individually" and "collaboratively" are meant strictly to describe group dynamics in each setting.

Despite analogous participation modalities, learning and problem-solving literature should not be directly compared without exploring how they are related. Situated learning offers one framework for understanding this relationship. Proposed by Lave and Wenger in the early 1990's, situated learning broadly suggests that learning cannot, and should not, be separated from the environment in which it is to be later applied (Lave & Wenger, 1991). Instead, what characterizes participation in "real-world" groups like those found in a workplace setting - where boundaries of engagement are not always well-defined and individual actors immerse themselves in community activities to varying extents - may also characterize participation in learning settings. From this arises the notion of "legitimate peripheral participation" (LPP), which suggests that over time, learners may increase their community engagement as they grow accustomed to, and help shape, the norms and expectations of the broader group (Lave & Wenger, 1991). LPP is one aspect of what Lave and Wenger call "Communities of Practice" (CoP), or groups that of individuals that share a common profession or trade (Lave, 1991) and engage in collaboration to achieve a set of shared objectives.

The concept of a community of practice in recent years has been broadly interpreted, adapted, and applied to better understand how people learn with each other online (Esteves, et al., 2008; Thomas, 2005). An outstanding question, however, is if the notion of a CoP - a structure that is characterized by increased engagement and collaboration among its participants over time - can apply to discussions in a massively open, online learning setting. In the past decade, some scholars have noticed the potential conflict between previously-held frameworks (like CoP) for describing and understanding learning in light of today's society - a society defined, in part, by interconnected webs of knowledge and information, many of which can be accessed instantly, on-demand. Connectivism, then, has emerged as an alternative learning theory that emphasizes the role of the individual as an actor in a larger, imperfect, uncertain, and sometimes chaotic web of knowledge-sharing[8] (Siemens, 2005).

It is too early in the history of MOOCs to determine if discussion participants should be considered as members of crowds or communities, but the literature offers strong foundation for exploring such issues. If we recall the majority demographic makeup of course participants so far – in some cases, over 80% with post-secondary degrees of some sort and primarily from the developed world (Belanger & Thornton, 2013) - we can begin to make inferences. MOOCs tend to draw people together based on unique areas of interest and shared professional pursuits (Kizilcec, et al., 2013; Belanger & Thornton, 2013). Because participants are lifelong learners coming from previous "real-world" jobs, families, and other experiences, theories like situated learning and situated cognition (Brown, et al., 1989) - in their emphasis on practical, real-world learning - can help explain how these individuals learn together. Their unique backgrounds join to create a tapestry of knowledge and insights where each individual has the potential to engage others through the information he/she disseminates and internalizes. Any subsequent knowledge generation is often a product of a chaotic and difficult-to-map process, just like in some professional and other real-world settings. In understanding the nature of MOOC participants' interactions - indeed, if they tend to cooperate or collaborate as evidenced by the discussion forums - it is important to look at how these conversations unfold around topics that relate to their lives outside of an educational setting.

---

[8] George Siemens has been a pioneer of MOOCs along with Dave Cormier, and most recently, has started the MOOC Research Initiative through support from the Bill and Melinda Gates Foundation.

Social network analysis: a theoretical framework for analysing communication

CSCL, situated learning, and connectivism together offer ideas that highlight a number of questions that seed this research: How engaged are participants in real-world learning settings? How do they tend to interact? To what extent do these learners cooperate and/or collaborate? To begin to answer these, however, a separate framework is necessary -- one that sheds light on the theory of how people can be "connected" to, and exchange information with, one another.

Network analysis has exploded in recent years as a method of investigating how individual actors in virtually all settings - friend groups, wireless components, microbiology - interact with one another (Easley & Kleinberg, 2010). A network, in the most basic sense, represents these actors as nodes that are connected by edges. These edges can be directed or undirected and are defined between nodes according to the application at hand (Psorakis, et al., 2012). For example, we can imagine a food transportation network where nodes represent pickup and drop off points for food, and edges represent the roads upon which freight carriers travel. The growth in network analysis research and applications has partially been a function of rapidly accelerating technological capacity as hardware and software improvements have enabled the creation and use of powerful visualization tools. However, it can also be attributed to a growing sense of interdependence and connectedness that has been brought about by globalization and the internet (Buchanan & Calderalli, 2010). From a pedagogical perspective, networks can serve as a framework for elucidating how individuals interact with one another in learning settings.

Social Network Analysis (SNA) is a particular realization of this network-driven approach. It has gained traction over the past two decades, however, as not simply a methodology, but also as a broader theoretical approach that informs how we can better understand social interactions. As Haythornthwaite suggests, SNA is predicated on the notion that individuals interact and exchange resources such as information. Empiricism is a key part of SNA: conclusions about communication between social actors can only be made after observing their interaction patterns and not *a priori* based on previous classifications. In this way, a collection of actors can only be described as belonging to a "group" once the network underlying their interactions is derived and interpreted (Haythornthwaite, 1996). Observing the network is a precursor, and fundamental, to understanding the existence of communities in social settings.

There are a number of studies that have applied the theory and techniques of SNA to technology-inclusive education settings. One study used SNA to better understand how 28 grade 5/6 school children interact with one another in a computer-supported collaborative environment (Palonen & Hakkarainen, 2000). Another paper recounted the ways in which 31 learners from engineering programs at two US-based universities collaborated online to design aerospace systems. In their analysis, the researchers combined a qualitative survey with their structural network analysis to gain a fuller picture of how students communicated, as well as the communication styles of each individual (Cho, et al., 2007). Many of these studies have built upon previous literature on distributed learning via computer mediated communication (CMC), some of which argues that communications online - with their capacity to inject emotion via emoticons and other methods - may not be as socially limiting as perceived by proponents of offline, in-person alternatives (Haythornthwaite, 2002).

The rise of "big data"[9] (boyd & Crawford, 2012) in recent years and the tools that enable its analysis have encouraged more recent studies to leverage the theory and practice of SNA in learning

---

[9] "Big data" has emerged as an overarching term that describes the complex datasets being generated in numerous fields. Data can be "big" for a variety of reasons, for example, because of its volume, granularity, or perhaps even the impact it has on a broader questions or investigation area.

contexts at unprecedented scales. In a 2006 *Science* report, researchers discussed how data from over 50,000 time-stamped email exchanges between university staff and students at a large university could help indicate the likelihood of future relationship formation (Kossinets & Watts, 2006). A 2013 *Nature* study used SNA as a guiding framework for asserting that the formation of "rich clubs" - or tight-knit learning communities created in the first few weeks of class - characterized high-performing students' behaviour, whereas low-performing ones tended to have inconsistent interactions with other low-performing students at later intervals. Their dataset consisted of fine-grained file-transfer and communication traces from 290 students using a virtual learning environment at a university in Spain, drawn over two consecutive years (Vaquero & Cebrian, 2013).

SNA and CSCL, then, go hand in hand. SNA provides a practical, observation-driven framework for understanding how people communicate and learn in computer-mediated settings. This empiricism is appropriate because it accounts for the nuances of each individual learning environment -- it requires the researcher to re-think *a priori* assumptions as guided by what a given network shows. SNA can help identify the types of communication that indicate student performance in applied learning settings (Baldwin, et al., 1997), and perhaps even pinpoint how the positions of individual learners within networks relate to, and impact, how they engage with educational content (Vaquero & Cebrian, 2013). In some cases, however, the link between participation in discussion networks and final course performance is not always clear (Davies & Graff, 2005). With SNA applied to conversations in discussion forums, researchers can begin to uncover relationships between communication and learner performance in massively open online courses.

Research questions

The overarching objective of this study is to better understand who tends to use discussion forums in massively open online courses that engage students in real-world learning, and how the interactions between learners differs throughout the online discussion forums. This motivating objective can be further explored through the following research questions:

1. What are the characteristics of students that participate in MOOC discussion forums?
2. What topics do students tend to discuss with one another and how do these discussions evolve over time and across topics?
3. To what extent do crowds and communities emerge in the discussion settings?
4. How does participation in discussion forums relate to students' performance?

Social constructivism and activity theory suggested that the cultural backgrounds and demographics of learners affect how they interact with – and therefore learn from – one another, justifying research question (1). Situated learning theory offers a framework for investigating what people tend to discuss in learning settings – hence, asking (2) in the specific context of MOOCs. The theory of SNA, principals of Connectivism and CSCL's distinction between collaboration and cooperation together justify (3) by calling for a more nuanced exploration of which students interact with which other students – and how SNA's tools can be used to better understand this communication. Finally, Question (4) is important as it enables this work to be interpreted in light of existing literature on learner interactions and achievement (e.g., Vaquero & Cebrian, 2013) by connecting the nature of students' interactions to one type of academic outcome.

Together, these questions aim to contribute to a broader understanding of how discussion forums can be utilized to enhance learning. They draw upon traditional learning theories to first

understand how students engage in discussions, and augment them with insights and tools from more recent frameworks in education and sociology. The data generated by discussion forum participation in MOOCs is undoubtedly "big" and therefore calls for a methodology that leverages both qualitative and quantitative tools in order to understand students' interactive behaviours in these settings.

Methodology

Real-world discussions in a MOOC: A case study

This investigation takes a case study approach (Bassey, 1999) in order to gain insights into how people use MOOCs to communicate with one another. Given the current early stage of MOOC research, it was decided that a broad analysis of forum participation in a specific context - instead of a randomized control trial with a specific hypotheses tested through manipulation of control and treatment groups - would be a more appropriate approach for this work.

Before discussing methodology, it is important to first understand the nature of the course selected for the case study and why it was chosen.

Course basics

The subject of the case study is a MOOC titled "Foundations of Business Strategy"[10], offered by Professor Michael Lenox[11] through the University of Virginia's Darden School of Business. The six-week long course started on March 4th and was offered via the Coursera MOOC platform. The goal of the course was to introduce the concept of strategic analysis in order to equip students to think critically about how they can start or improve businesses across industries. It utilized frameworks that are widely used in the business world. It also leveraged tangible examples of companies operating strategically in their respective markets. These examples helped create an applied, real-world learning environment.

Given the focus of this study, the course's real-world relevance made it an ideal partner for further investigation. The course was selected for its emphasis on practical applications of course material, since it was anticipated that this would attract a student body with diverse professional experiences from around the world. Moreover, the content was delivered by a well-known professor in the field by a reputable business school, perhaps contributing to the large enrolment numbers. Finally, its use of business school pedagogical practices, such as business case discussions, encouraged interactions among the diverse range of students and yielded an opportunity to explore these interactions through research.

Each week, students were responsible for watching a series of short, 5-10 minute lecture segments (totalling approximately 1-2 hours per week) and reading segments from the course's underlying text, *The Strategist's Toolkit[12]*. There were mandatory weekly quizzes that aimed at reinforcing lecture content. Students were also encouraged to read a business case study each week and to discuss the cases in the discussion forums by answer a series of related questions.

---

[10] The course can be found here, and will be offered again in September 2013: https://www.coursera.org/course/strategy101

[11] The instructor has agreed to leave himself and the course as not-anonymized for the purposes of this research. Full disclosure of identity will enable future work to build more thoroughly upon the findings presented in this work. This is in line with CUREC approval and Coursera's guidelines.

[12] http://blogs.darden.virginia.edu/strategiststoolkit/

Almost 87,000 students enrolled in the course. Of those, just over half logged in at least once and were considered "active participants". Enrolment increased rapidly as the course's start date grew closer: 12,000 people were registered by late October, followed by 52,000 in early February. Nearly 10,000 students joined after the course began on March 4th.

Final project

In addition to the lectures, readings, and case discussions, students were required to complete a final project. The project asked students to perform a strategic analysis of an organization of their choosing. In order to further emphasize real-world learning and problem solving, I worked with the professor to invite small enterprises, non-profits, and other organizations to join the course and solicit the strategic insights of students. For their final projects, students could perform a "live" cases analysis of one of these organizations, although they were not required to do so and could still select an organization that had not been recruited for this purpose. Their final analyses were evaluated through peer-assessment. A press release was distributed to raise awareness among both students and organizations about the real-world problem solving opportunity[13].

The discussion forum

The course's discussion forum was comprised of the following sub-forums: Technical feedback, where students offered suggestions for improving the technological components of the course; Course material feedback, where students' suggestions focused more on how to improve content and the communication of key information; Study groups, where students were able to self-organize and arrange meet-ups in their respective localities or online; Readings, where students could discuss concepts covered in the text; Lectures, which facilitated discussion around topics covered in lecture videos; Cases, which engaged students in weekly discussions about the strategic decisions and market position of companies such as Google, Apple, and Disney; and Final projects, which enabled students to ask questions and engage others on the final case analysis assignment.

In the Cases sub-forum, additional sub-forums were created for each case, which in turn had sub-forums that were dedicated to asking specific questions that helped facilitate student participation. Moreover, one week after the course commenced, a sub-forum was added to the Final projects sub-forum to enable organizations to post information about their companies in order to recruit students' assistance. As evidenced by their topics and overall structure, the discussion forum and sub-forums were designed to promote authentic interactions and mutual exploration of real-world business strategy issues.

---

[13] http://www.darden.virginia.edu/web/Media/Darden-News-Articles/2013/UVa-Darden-Strategy-MOOC-Enables-Students-to-Help-Entrepreneurs-and-Nonprofits/

	Sub-forum	Latest Activity
Discussion Forums	**General Discussion** General discussion about the course, life, and everything under the sun.	Anyone interested in white papers on... (an hour ago)
Programming Assignments	**Data Science Examples** Post you favorite data science examples from around the web -- even better if you made it yourself!	Fuzzy VLOOKUP in excel (6 days ago)
Quizzes	**Study Groups** Find friends and arrange meet ups!	Moscow Post-Course Meetup 14 July 17:00 (a day ago)
Peer Assessments	**Lectures** Specific questions about the lectures.	Video Lecture copyrights (a day ago)
Optional Real-World Project	**Assignments** Specific questions and clarifications about the assignments.	What we learnt from the Tableau... (21 hours ago)
Syllabus		
Course Logistics	**Real-World Projects** A chance for you to discuss real-world data science projects!	How do countries deal with employment... (a day ago)
Running Tableau on AWS	**Course Material Feedback** Please use this forum to report any potential errors in the lectures, assignments/grading, and other course materials to the teaching staff.	Wording issue in assignment 1 (3 days ago)
Class Virtual Machine	**Technical Feedback** Please report any video playback issues, 404 errors, and other technical issues and bugs to Coursera staff here.	How about using R instead of Tableau? (13 days ago)
Github Instructions		

Figure 1: An example view of the forum structure of one particular Coursera-based MOOC.

Primary methods of investigation

The course selected for this case study offered a unique research opportunity. Sometimes, when data is analysed post-hoc, the research is not always aware of the learners' contexts and therefore may be unable to conduct a robust investigation. Understanding this context is an important part of doing education research (Barab, 2004). I worked with the University of Virginia's Darden School of Business help design part of the course (namely, the final project). I also helped administer pre and post-course surveys in addition to analysing learners' "digital traces" and recording informal observations of discussion forum participation.

Each research method - surveys, observation and the capture of all digital trace data - offers unique insights into students' experiences, and more relevant to this study, how participants tended to engage with each other in the discussion forums. The surveys extracted qualitative information about learners' experiences in the course but also offered key demographic insights that aided the investigation of forum discussions. Digital trace data enabled forum participation to be quantified and analysed in the context of other variables, such as geography and final marks. Observations supplemented the quantitative insights of the aforementioned approaches with specific anecdotes from forum use.

The following subsections delve into the nuances of each data source and discuss the tools and methods used to analyse them collectively as part of a unified framework.

Pre and post-course surveys

In order to gain insights into course participants' backgrounds, motivations, and overall experiences with the course, both pre and post-surveys were administered through the MOOC platform. As noted above, the course invited real organizations to join the course and connect with students to solicit business strategy insights. With these two (likely overlapping) groups of participants in the course – students and organizations - two separate surveys were administered during both the start and end of the course: one for those individuals that identified as being part of an "organization"[14], and another for those that identified as being "students." Surveying both of these groups offered deeper insights into the forum participation of learners by exploring the

---

[14] While this was originally intended to mean only those individuals that registered for the course to recruit the help of students, many individuals that registered as students but also identified as being part of an "organization" ended up filling out this survey.

natural relationships between their background characteristics and their communication tendencies (Punch, 2003).

The pre-course surveys were administered online during the first week of the course. Professor Lenox used a class-wide announcement to encourage both organizations and students to participate. The organization and student pre-course surveys had 7 and 8 questions, respectively. Both asked about participants' previous educational experiences, current professions, and motivations for joining the course. These surveys aided in exploring research question (1) by providing insights into the backgrounds and demographics of a subset of the course's registrants, and more specific to this study, forum participants.

In addition to repeating the questions on the pre-course survey, the post-course questionnaires had additional questions that gauged how much time participants spent on the course and each individual's overall experience. Both post-course surveys had 26 questions and were distributed through the online course interface. Using the course platform to manage survey creation and distribution instead of a third-party solution was a critical part of this research, as it enabled survey responses to be linked to the digital traces collected for each participant. Copies of the pre and post-course surveys for both groups are included in the Appendix.

Over 1,800 organizations and 7,000 students responded to the pre-course survey. The 7,000 responses represented approximately 9% of all enrolled students at the start of the course, and almost 15% of those that logged in at least once. Post-course responses were less voluminous: approximately 1,100 organizations and 1,000 students completed these questionnaires. Despite the large number of responses to both sets of surveys, these results must be caveated with the likely existence of response bias (Armstrong & Overton, 1977; Couper, 2000). Most respondents were those that completed the course, even though a large proportion of registrants never once logged in. This suggests that survey respondents tended to be learners that were committed to earning certification for their participation in the MOOC. Even with large sample sizes, it is therefore not possible to claim that the responses were representative of the thoughts and sentiments of the course's student body (Hedges & Olkin, 1987).

As more MOOCs are offered, it will be important to find alternative ways of collecting demographic and motivation information beyond surveys in order to construct a more accurate picture of students' learning in these settings.

Learners' digital traces

The cornerstone of this study's methodology is the analysis of granular learner data. This data was recorded throughout the course in a MySQL relational database and exported in three main files: The full-text of assignment submissions, forum discussion participation (including the full-text of each forum post and comments), and other general course participation information (including information on how students fared in completing peer assessments, as well as their final course marks). Responses to the aforementioned surveys were also included in the dataset, which enabled data collected on participants to be compared alongside self-reported measures. The identity of each student was protected through the use of anonymised user IDs. In all, records for over 87,000 participants were made available through 50 database tables, with each table containing anywhere from 2 to 20 fields.

The temporal granularity of this data is also immense: each assignment submission, lecture view/download, and forum post/comment/thread creation was time-stamped. To narrow scope and focus on big-picture trends, only data captured in the relational database was analysed as a part of this study, although clickstream data that recorded even more fine-grained user interactions with the platform may be analysed in a future project.

Beyond recent papers that have emerged on MOOCs (Breslow, et al., 2013; Kizilcec, et al., 2013), the scale of this dataset is unprecedented in previous education research. A major challenge in formulating the methodology for investigating this data included determining which data to analyse and to what extent[15]. Given the study's emphasis on learning how participants interact through the course's discussion forums, a Social Network Analysis (SNA) approach was employed to visualize and interpret the nature of course-related communications. As noted in the literature review, SNA is used to depict how actors in some setting share resources with one another. These actors are represented by nodes in the network, and edges are used to connect nodes that are related in some way (Haythornthwaite, 1996).

In this specific context, SNA includes a broad analysis of user participation in online discussions and which discussion threads/forums tend to be viewed and posted to the most. The analysis also includes network visualizations to depict how discussions emerged around various topics. Visualizations are utilized largely because of the amount of information about discussions they can capture at once - including demographics on forum participants, the nature and flow of information in the forums, and each participant's contribution volume, among other factors (Easley & Kleinberg, 2010). Because of time constraints, which were compounded by the scale of the data, formal content analysis methods (Krippendorff, 2013) were not employed, although future research will involve a more in-depth analysis of the communication.

The following subsections detail how learner data, particularly records of their forum interactions, was handled in order to conduct the investigation.

Technical preparations

An important precondition for handling a dataset of this size and scope was an efficient, scalable technical pipeline for extracting, manipulating, and interpreting key metrics. To mirror Coursera's database schema, a MySQL instance was created and populated using the files that comprised the initial data export. The dataset was manipulated using standard Python[16] libraries. The NetworkX[17] Python library was used to help define and visualize the communication networks. In all, over 3,000 lines of code were written as a part of this investigation. I hope to open-source these analytical tools in the future in order to support other researchers that wish to perform similar investigations on MOOC datasets.

Different network topologies for learners' forum discussions

The social networks that underpin learners' interactions in MOOC discussion forums are not obvious at first glance. Given this study's interest in understanding who uses discussion forums in MOOCs with real-world learning and how, it is reasonable to make nodes in the underlying network model represent individual participants. However, edges between these nodes can be defined in a number of ways as they have been in previous education research (Huerta-Quintanilla, et al., 2013). There are many logical topologies[18], or representations of how interactions and communications occur, that can be used to define and visualize social networks that represent MOOC forum discussions.

This study employs two different topological definitions. The first is based on a rather coarse metric: two nodes, i.e. two forum participants, are connected with an edge if they happen to

---

[15] This is covered more extensively in the Results and Discussion sections.

[16] A common programming language used for statistical and other analysis. See http://www.python.org/.

[17] Developed by Los Alamos National Labs in the USA to aid in network-based research. See http://networkx.github.io/.

[18] http://wordnetweb.princeton.edu/perl/webwn?s=logical%20topology

participate in at least one discussion forum thread together. Participation in this case is defined as posting text to the discussion forum. Simply viewing posts without adding to the discussion explicitly does not qualify as participation, although previous literature has focused on investigating this type of forum-lurking behaviour more explicitly (Preece, et al., 2004). These edges are undirected as there is no obvious way to discern which way information flows between actors in this relationship.

The second definition utilizes directed edges to describe forum post and forum comment relationships. A directed edge exists from person A to person B if A commented on at least one of B's forum posts. The directionality in these "post-comment networks" offers insights into the role each individual plays in the various discussions he/she is involved in (Wasserman & Faust, 1994).

The study considers both of these topologies for a number of reasons. The "thread networks" capture relationships that are based on shared interest and exposure: co-participation in a discussion thread, although a relatively weak tie, still implies that two participants were interested enough in the same topic to take part in discussion together. Still, the granularity of this interest is lost and it is unclear which *parts* of a discussion on a particular thread captured which subsets of individuals. "Post-comment networks" offer this granular insight by specifying how individuals interacted with each other on a particular forum thread. However, they are subject to nuances of the discussion platform as some participants opted not to comment on posts (and instead, simply created new forum posts) even when their new posts were, indeed, responses to earlier statements by their peers. Considering both topologies enables a more robust analysis across the forms and functions of participant interactions in the MOOC, and this was supported by the observations conducted as a part of this study.

Spatio-temporal network definitions

After deciding upon the network topologies to explore, the next step was to determine how the social network analysis would be segmented to help expose interesting communication trends. The first stage of data analysis included looking at social networks created for the entire 7-week period, encompassing all of the different sub-forums. However, with over 4,000 forum participants, this topology yielded a densely connected set of nodes that made it difficult to gain insights into the nuances of communication between the different learners. The Figure 2 shows the complete network:
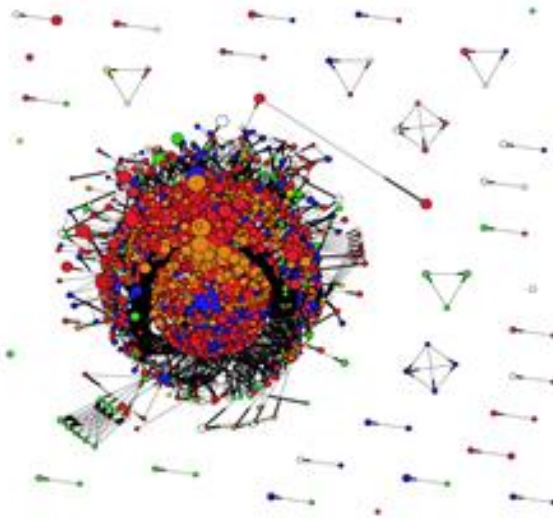
Figure 2: The network of all forum participants in the MOOC, connected based on thread participation.

To mitigate this "noisiness", networks were then created according to "space-time" delineations as they occurred in the course. For one, week-by-week snapshots of discussion forum activity were visualized by plotting only forum interactions as they occurred in a particular week. Since the course started on March 4th (a Monday), weeks were defined as Monday to Sunday. The membership of a forum post or comment in a particular week was determined based on that forum interaction's timestamp in the relational database. Additionally, networks were visualized according to how discussions evolved in each sub-forum. To achieve simultaneous spatial and temporal granularity, networks were also computed on a week-by-week basis for each of the different sub-forums in the course. Spatio-temporal windows for finer-grain social network analysis have been used in previous research across disciplines (Slingsby, et al., 2012).

To summarize, the structure of communication in the MOOC's discussion forums is captured in this study by two primary dimensions: the network topology used and its spatio-temporal decomposition. Table 1 depicts these dimensions, with x's to indicate which of these pairs were analysed in this study.

	Weekly snapshot	Forum snapshot	Weekly-forum snapshot
Thread network	x	x	x
Post-comment network	x	x	x

Table 1: The different discussion forum social networks computed as a part of this study.

Varying node attributes to capture additional information

In order to capture and reveal information about the discussion participants themselves, a third dimension was considered: participant attributes. One of these attributes was the level of each student's participation in the discussion forums. Node sizes in the networks were manipulated

according to the user's communication activity in that particular setting -- larger nodes implied more active use[19]. Node colour was used to depict a number of foundational attributes, namely: highest level of prior education, primary motivation for taking the course and geography. The first two characteristics were extracted from pre-course survey responses; geography was determined by using a web service to connect the last-accessed IP addressed recorded for each user to a specific country[20]. To maintain ethical best practices in this research, none of the published results identify the exact country of any given forum participant. Nodes in the network visualizations were also coloured according to participant's final marks in the course. The next section offers a more granular look at the results of these network visualizations and how they can enable a deeper understanding of communication.

Capturing this additional dimension - the attributes of forum participants - is critical to informing not just *how* discussions in MOOCs evolve, but also, *who* is responsible for evolving them, and how their participation relates to their overall performance in the course.

Informal observations

While learners' digital traces and survey responses enabled a broad understanding of how interactions evolved throughout the course, on their own, they offer an incomplete picture of how students communicated with one another. While this study was not an ethnography, it adopted tools from ethnographic research by integrating an analysis of learner interactions throughout the research process (Emerson, et al., 2001). Informal observations of forum activity were recorded through approximately 2-3 hours of weekly analysis during the course's progression. The objective of these observations was to look into specific interactions between students in the different sub-forums in order to begin to understand the type of content that was being generated -- largely, to add a qualitative layer to the data-driven network visualization aspect of the research. These observations helped explore the research questions by confirming the demographic trends observed through the surveys as well as the patterns of communication interpreted through the SNA component of the investigation. Notes on these observations were recorded on average every other day for seven weeks[21], organized by date and sub-forum and totalling over 10,000 words.

Results

Background information on course participants

It is important to first explore the rich diversity of backgrounds and experiences of the nearly 87,000 course registrants. The digital trace data revealed that approximately 143 countries were represented from six different continents (including Oceania). Over 2,500 students received a nonzero score in the course, implying that they submitted at least one quiz or the final project. Over 47,000 students were considered "active" (i.e. logged into the course's website at least once) and more than 39,000 of them viewed or downloaded at least one of the lecture segments. These

---

[19] A simple algorithm that assumed the following weight-prioritization (in descending order) was used: Thread creations, forum posts, forum post comments. The algorithm assigned points based on a given users' actions and a given node size was determined based on these point totals.

[20] This service enabled countries to be retrieved for the IP addresses stored in the relational database: http://ipinfodb.com/ip_location_api.php.

[21] The seventh week of the course was dedicated to students' peer-assessments of each others' final project submissions.

trends indicate the different reasons people appear to take MOOCs and reflect the different interpretations of what it means to "participate" in these settings (Kizilcec, et al., 2013).

Based on over 7,000 responses to the pre-course survey, participants were primarily well-educated young adults from developed countries with priori experience in industry. Nearly 40% of respondents were between the ages of 25 and 34, over 80% had at least a Bachelor's degree, and approximately 40% had a Master's or professional degree (e.g. MBA, J.D., etc.). Moreover, approximately 54% of pre-course and 58% of post-course survey respondents identified themselves as "industry professionals." The most popular motivation for taking the course was "professional development."

It is important to note the following characteristics of course participants: most were young professionals, had substantial previous educational experiences, and were seeking skills to help them advance in their careers. This tendency can be interpreted in the context of another feature: 67% of students in the pre-course survey and 64% in the post-course survey indicated that their peers played an important role in their learning processes. These participants were not the typical undergraduate cohort. Rather, many saw the course as an opportunity to engage with real-world business issues by connecting with other like-minded people from around the world.

Background attributes of forum users

From the survey responses, it is evident that many students wanted to gain tangible skills from the course – often, by interacting with their peers – that they could carry over to their lives outside of the MOOC. This desire was reflected in the discussion forums, where over 4,000 participants created at least one post or comment. This number is 60% higher than the aforementioned 2,500 individuals that received a nonzero final score, suggesting that a large number of forum users participated in discussions without submitting any of the required assignments. Forum participants accounted for just over 4% (4,000) of all course registrants. Still, while comprising a small proportion of all users, forum participants are still a large and highly interesting focus of study. The remainder of this analysis will focus on these individuals.

Participants' demographics and motivations

Understanding *who* used the forums in the MOOC was fundamental to gauging *how* this participation occurred. Geographic information could be identified for 3,631 out of 4,337 total forum participants[22]. 57% of identifiable forum participants were from either Europe or North America and 24% were from Asia. Table 2 shows the proportion of forum participants by continent. For comparison purposes, the course-wide proportions for those with known location information are also shown.

Continent	% Forum participants in this continent (N)	% Course participants in this continent (N)
North America	32% (1,145)	32% (15,481)
South America	10% (372)	7% (3,496)
Europe	25% (918)	28% (13,412)
Asia	24% (879)	26% (12,723)
Africa	7% (234)	5% (2,453)
Oceania	2% (83)	2% (1,125)

---

[22] A failure to capture geographic information for users is a result of the sparseness of some parts of the database, most likely caused by errors in logging information as users accessed the online platform.

1,984 forum participants indicated their highest level of prior education on the pre-course surveys. 83% of these individuals had a Bachelor's or Master's/Professional degree. Approximately 10% of forum participants for whom data was available indicated "Some College" as their highest level of educational attainment. Table 3 shows the educational attainment for forum participants and course registrants.

Highest level of educational attainment	% Forum participants with this level (N)	% Course participants with this level (N)
Some High School	1% (26)	1% (103)
Completed High School	3% (66)	4% (282)
Some College	10% (204)	11% (805)
Bachelor's Degree	43% (849)	42% (3,050)
Master's/Professional Degree	40% (781)	39% (2,832)
Doctorate	3% (58)	4% (265)

Table 3: Proportion of forum users and total course participations (based on pre-course survey completion) and their highest level of educational attainment.

Based on the 1,964 forum participants that indicated their motivation[23] for joining the course on the pre-course survey, nearly 93% indicated professional development as their primary motivation; 6% participated mainly out of curiosity and a desire for enjoyment; 1% participated primarily because they didn't have access to other educational resources, and less than 1% because the course was offered by a prestigious institution. The course-wide participation numbers were virtually identical. This suggests that students taking the course were primarily interested in gaining skills and knowledge that they could then apply to their professional lives – and that the forum participants, in terms of motivation, were indistinguishable from other course participants.

Trends underlying forum participation

The data above offers some initial insights into the background attributes of forum users. The next section offers insights into some of the specific trends that characterized their interactions.

Per-user trends

While textual analysis was not performed as a part of this study due to time constraints, I computed the average length (number of words) of each forum post and comment. The average

---

[23] The survey question about motivation enabled multiple responses. A simple "majority voting" algorithm was therefore implemented to discern a user's primary motivation for taking the course. Voting ties were broken with the following prioritization: Access to other resources, Professional Development, Curiosity and/or Enjoyment, and Prestige of institution. So, for example, if an individual indicated a lack of access to other resources and professional development as motivations for taking the course, her primary motivation would be labeled as a lack of access to other resources.

length of each user's posts and comments is a coarse indication of the extent of each instance of forum participation. Figure 3 show the trends across all forum threads.
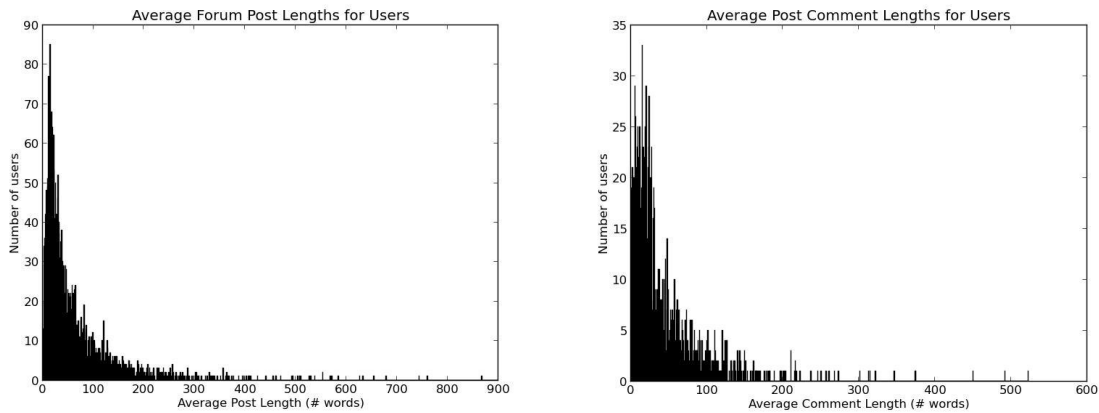

Figure 3: Average length (in number of words) of users' forum posts and forum comments.

Here, the x-axis represents the average number of words comprising each post or comment and the y-axis indicates the number of users that reflected this posting behaviour. Upon inspection, both histograms depict a lognormal-like distribution (Limpert, et al., 2001) over the average size of posts and comments. The long-tailed distribution implies that while most students had posts and/or comments that were below 50 words in length, there were a few participants that tended to post more detailed content.

Trends in users' thread creations, forum posts, and forum comments all mirrored one another. In all three cases, a large proportion of users tended to engage in small amounts of activity. The histograms in Figure 4 depict this activity in more detail. There appears to be an exponential decline in the number of users that engage in incrementally larger and larger amounts of forum activity. This is not surprising given the different ways in which MOOC participants tend to engage with courses (Kizilcec, et al., 2013). Many students that start off completing assignments and watching lectures eventually stop participating as they lose interest or can no longer keep pace with course content. These sort of exponential declines are characteristic of many other "scale-free" social networks, suggesting the highly social nature of MOOC discussion forums (Barabási, 2009).
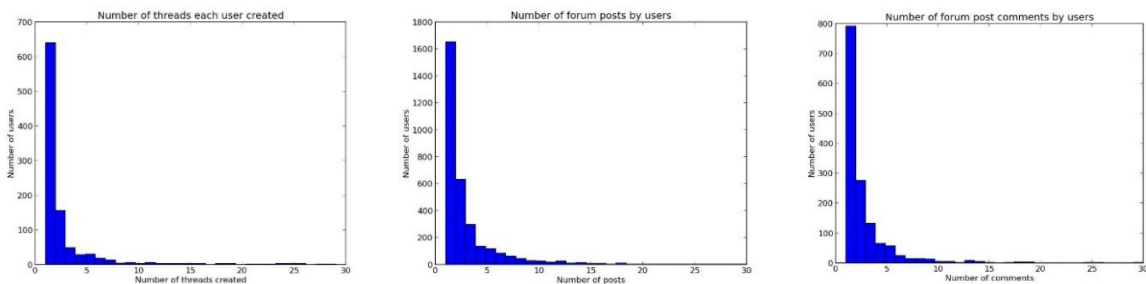

Figure 4: (from left to right) – Distribution of the number of thread creations, forum posts, and forum comments per forum participant.

Forum reputation points were also available per user. On Coursera, a user's forum reputation is determined by the sumtotal of upvotes and downvotes on all of her posts as inputted by her classmates. Figure 5 reveals that a large proportion of students had a positive forum rating, with most receiving at least 1 upvote in the aggregate. While this is a coarse metric that does not capture the nuances of discussion sentiments, it does indicate that the forums encouraged positive feedback and cultivated a generally supportive environment.
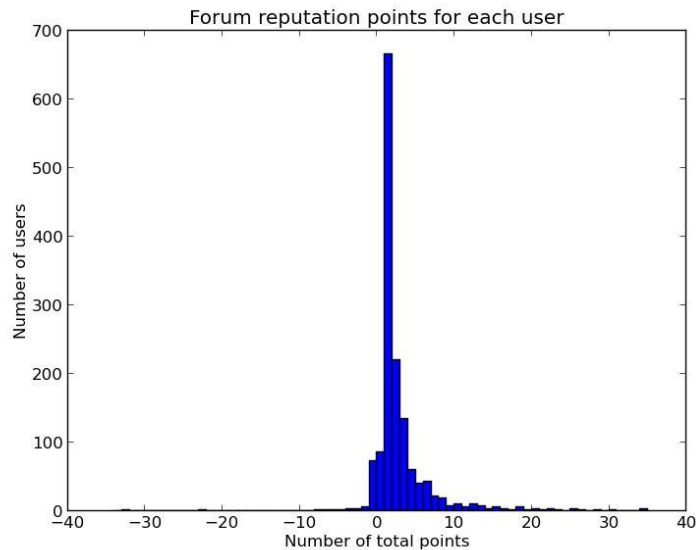
Figure 5: Total forum reputation points per user, computed as the sum of upvotes and downvotes.

Popular sub-forums and discussion threads

I used the titles of sub-forums and discussions threads to discern different discussion topic preferences across users. Specifically, I aimed to learn which sub-forums and threads were most popular. Guided by the metrics available in the relational database, I used the number of views and number of posts per sub-forum and thread to measure popularity.

Tables 4 and 5 show the top 5 most popular sub-forums and discussion threads according to the number of views and posts, respectively. In both sub-forums and threads, the highest number of views were for topics related to the final project. Some of the most popular sub-forums were also related to case analyses from earlier parts of the course, namely, questions that guided students through an analysis of Google's business strategy. In addition to discussing case analyses and the final project, students also used the forums in order to form study groups with one another. These study groups were often created with respect to students' geographies. Informal observations revealed that some of the forum posts in study group threads for particular countries were written in languages other than English.

This data sheds light on a key characteristic of communication in this particular MOOC: many participants used the forums to seek out connections with their peers and discuss how business strategy tools could be used to address real-world challenges.

Top 5 Sub-forums (# of views)	Top 5 threads (# of views)
Final Project (40179)	Sample project report (4886)
Study Groups (24567)	Confused on final project (3379)
Week 1: Strategists Toolkit: Introduction / Competitor Analysis / Environmental Analysis (23579)	The problem with identifying competitors (2543)
Review Google's corporate webpage and annual report. How would you describe Google's strategy? (20089)	European Time Zone - Forming a Group (2538)
Week 1: Introduction to Strategic Analysis (11034)	India Study Group (2345)

Table 4: Top 5 sub-forums and threads with the highest number of views.

Top 5 Sub-forums (# of posts)	Top 5 threads (# of posts)
Study Groups (1933)	INDIA study group (132 posts)
Final Project (1564)	Spanish Study Group (127)
Week 1: Strategists Toolkit: Introduction / Competitor Analysis / Environmental Analysis (1465)	Asian Timezone Group (127)
Review Google's corporate webpage and annual report. How would you describe Google's strategy? (815)	European Time Zone - Forming a Group (124)
Week 1: Introduction to Strategic Analysis (513)	Why are you taking this class? (108)

Table 5: Top 5 sub-forums and threads with the highest number of posts.

Social network analysis: visualizing virtual discussions about real-world topics

From the data so far, it is evident that most users participated very little in the forums, and that the most popular topics of conversation tended to be focused on discussions with practical, real-world relevance. These measures on their own, however, do not offer a nuanced picture of *how* discussions between individuals occurred during the course. For this, we turn to social network analysis, namely, the computation and visualization of the different communication networks underlying the forums.

As presented in the methodology, Table 6 depicts the different networks considered as a part of this work. Here, rows denote how edges in the network were defined: "Thread networks" placed an undirected edge between nodes (forum users) A and B if both participated in the same discussion thread. "Post-comment networks" defined a directed edge from A to B if A commented on at least one of B's forum posts. The columns indicate the spatio-temporal network "snapshots" that were visualized. Therefore, the entry with "row,column" values equal to "Thread-network, Weekly snapshot" refers to the set of network visualizations that depict which users participated in the same threads as others during each week of the course. Both sets of network visualizations

used a spring layout algorithm via the NetworkX library, which aimed to minimize the number of overlapping edges[24].

	Weekly snapshot	Forum snapshot	Weekly-forum snapshot
Thread network	x	X	x
Post-comment network	x	X	x

Table 6: The different network topologies computed as a part of this study.

The networks resulting from this computation were analysed in terms of both qualitative and quantitative attributes. Leveraging this hybrid toolkit proved invaluable in gaining a more thorough understanding of how discussions around real-world topics emerged and evolved. The colours used for nodes in the networks represent a particular attribute of forum participants. To preserve confidentiality of participants, the represented attributes will not be discussed in relation to the network diagrams themselves, and instead, only in aggregate terms in the analysis that follows.

Key structural trends observed via thread networks

Thread network visualizations tended to be dense[25] because the criteria for connecting two nodes was relatively weak (Figure 6 shows an example of the network defined over the entire discussion forum). Therefore, the most granular snapshots were the weekly-forum visualizations because they showed how discussions occurred in each sub-forum, segmented by week. While some of these visualizations were still dense, they offered more detail and insights into the nuances of forum discussion patterns than the other snapshots. Thus, part of the analysis included computing the thread network visualizations across the aforementioned spatio-temporal windows and selecting the most interesting ones for further analysis – where "interesting" was measured by the ability to discern distinct patterns of engagement between different participants. Below, we explore the thread-networks for sub-forums that exhibited particularly interesting patterns of activity.

---

[24] For more information on the spring layout algorithm, please refer to this resource: http://networkx.github.io/documentation/latest/reference/generated/networkx.drawing.layout.spring_layout.html

[25] Denseness in social networks means that the number of edges in the network is close to the number of total possible edges. See this resource for more information: http://wiki.gephi.org/index.php/Graph_Density
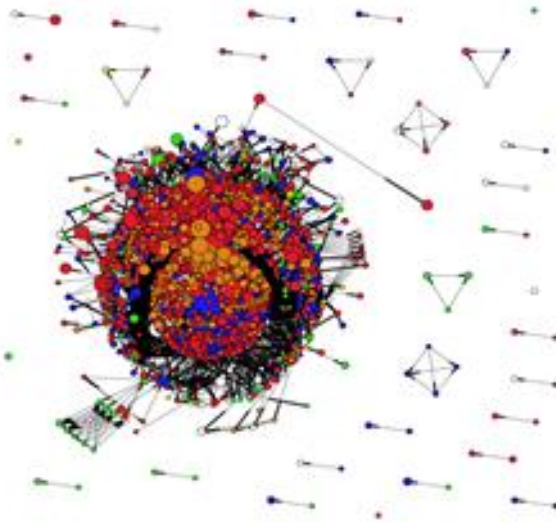
Figure 6: A dense network representing thread-interactions over the entire duration of the course.

Readings sub-forum

The Readings sub-forum enabled students to discuss the weekly textbook readings (not including case readings). Figure 7 captures the dramatic change in this sub-forum's communication trends from week 1 to week 2. Week 1 was characterized by a very dense network with a defined centre of communication. Most participants appeared to engage in a number of different threads; this yielded a network where many nodes are connected to many other nodes. In week 2, however, almost the opposite was true. Participation levels dropped significantly and forum users appeared to be more focused in their communications, participating in a small number of different discussions. The remaining weeks harboured similarly declining participation and smaller group-focused discussions became more typical.
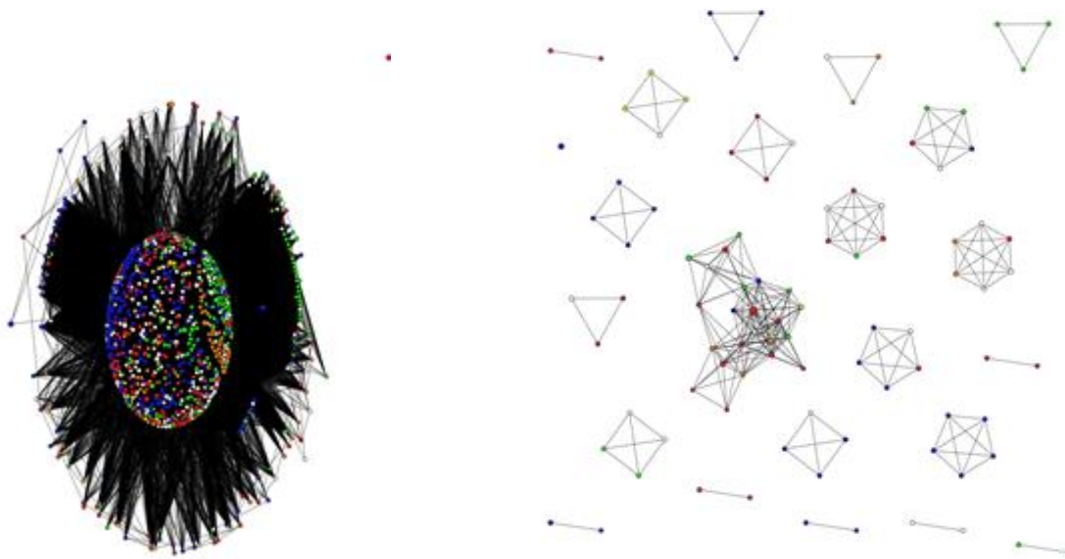
Figure 7: Thread participation in weeks 1 and 2 (left to right) of the Readings sub-forum.

Study Groups sub-forum

The Study Groups sub-forum enabled participants to connect with one another to form support networks around the course's content. Observations revealed that most threads in this sub-forum were created by students in order to form study groups around particular time zones, languages, or geographic locations. Even though the sub-forum was established to encourage the formation of small groups, week 1 showed a dense centre of communication with just a few isolated node groups on the periphery. This suggests that most students were not exclusive in their interactions and instead participated in many threads early on. Week 2 demonstrated a shift towards smaller groups of more focused, exclusive interaction. Participation dwindled in subsequent weeks and, as a result, the number of distinct study groups using the forums also decreased. Figure 8 shows how interactions in this particular network evolved throughout the course.
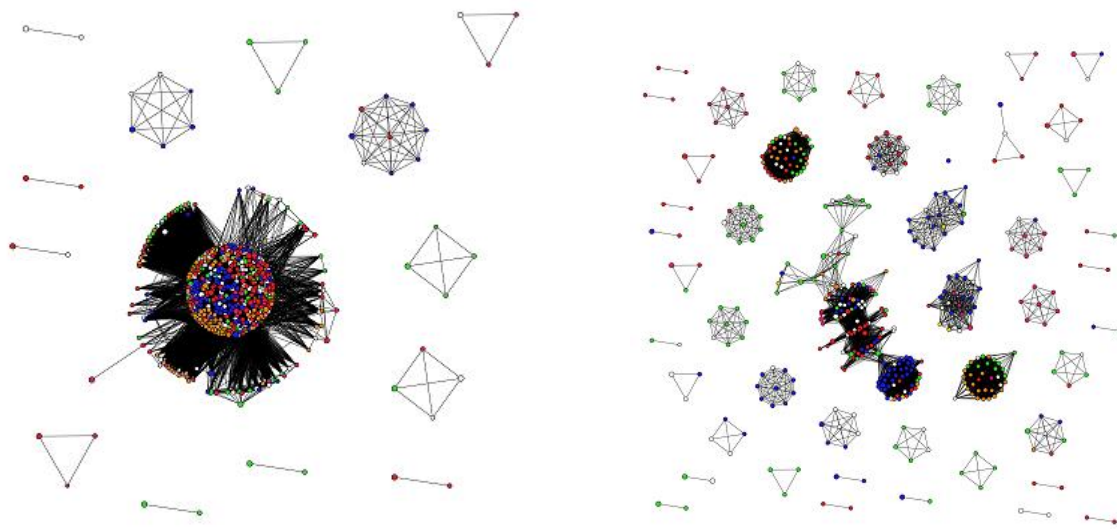
Figure 8: Thread participation in weeks 1 and 2 (left to right) of the Study Groups sub-forum.

Final Projects sub-forum

The course's final project asked students to perform a strategic analysis of any organization. The Final Projects sub-forum encouraged students to discuss questions and topics as they related to this assignment. Similar to the study groups and readings forums, week 1 was characterized by a dense network where most students engage in a number of different threads. Week 2 witnessed participation fragment into a number of smaller groups. These group structures persisted and participation dwindled as the course progressed. However, in the sixth week of the course - when the final project was due - participation increased dramatically and a dense centre of communication once again emerged. This trend persisted into week 7, when students were responsible for assessing one another's final analyses. The participation resurgence in week six appeared to result partly from questions about what the final analysis write-ups should entail. In week 7, much discussion focused on the mechanics of peer assessment, as well as how to handle cases where students had plagiarized their submissions. Figure 9 shows this evolution week-by-week.
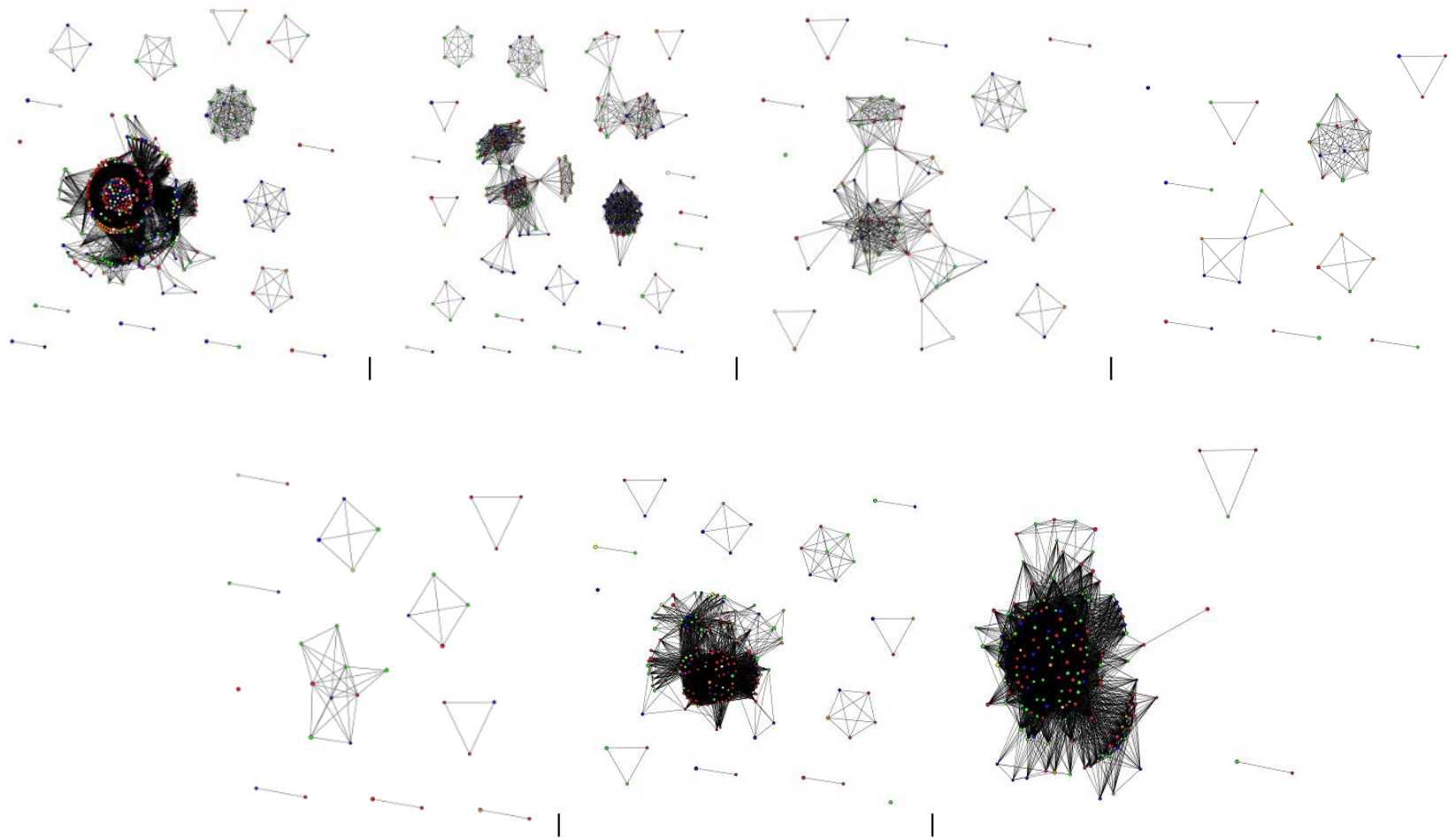
Figure 9: Thread participation in the Final Projects sub-forum for weeks 1 through 7 (left to right) of the course.

In order to further encourage real-world learning, the instructor invited small enterprises, non-profits, and other organizations to solicit the strategic insights of students. The Organizations Seeking Analysis sub-forum was contained within the final projects sub-forum; most threads within this sub-forum were created by organizations that used them to articulate one or more of their business strategy needs. Students that wanted to collaborate with these organizations expressed interest by posting or commenting on the appropriate threads. The sub-forum was opened at the end of the course's first week.

The network visualizations reveal similar trends to the previously-discussed sub-forums: dwindling participation and group-fragmentation into more exclusive discussions as the course progressed. The sub-forum's first week did not have as dense of a communication centre as some of the other sub-forums, but it harboured more communication diversity than subsequent weeks. Through the informal observations of these threads it became clear that diminishing participation was likely due to "leakage" to other platforms: students tended to create Facebook groups or relied on email and other media to continue their communications. This trend was appropriate for this sub-forum as once students assembled into groups around particular real-world issue areas, there was no need to continue these discussions on the MOOC platform itself. Figure 10 illustrates some of this behaviour.
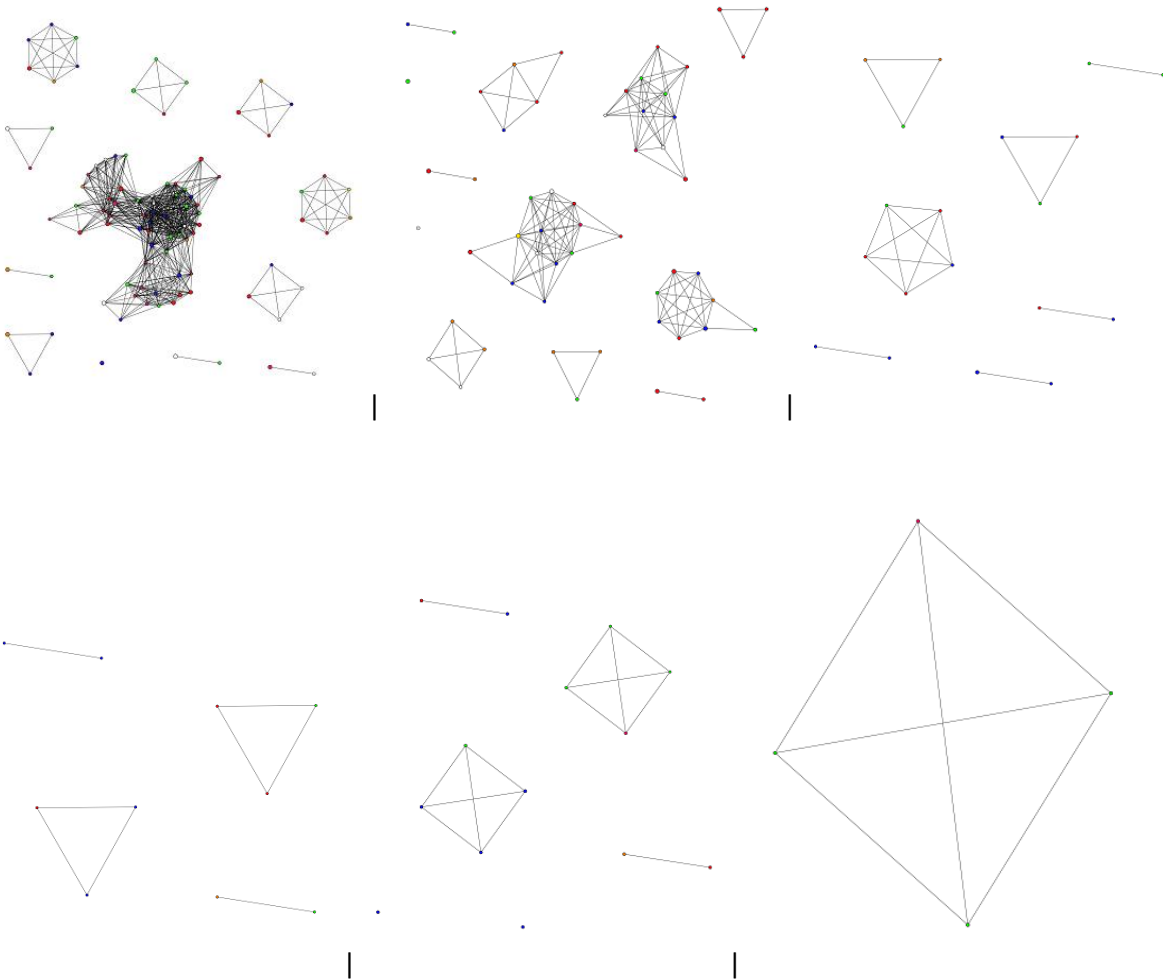
Figure 10: Thread participation in the Organizations Seeking Analysis sub-forum for weeks 2 through 7 (left to right) of the course.[26]

Cases sub-forum

Figure 11 shows discussions in the Cases sub-forum during the first, third, and sixth weeks of the course. As noted earlier, the Cases sub-forum was created to enable students to discuss and debate a series of case analysis questions put forth as a way to promote critical thinking around real-world business issues. From observation, we can see that the network holds a strong centre of communication throughout the course. The number of connections gradually falls, largely because participation fell as the course progressed. Moreover, we can see that as the weeks ensued, while a strong communication core persisted, there appeared to be more "isolated" discussions occurring - i.e., students began to focus on participating in just a few threads instead of many. During each week, there were some participants that engaged solely with a small group of other participants -- they remained entirely disconnected from other discussions that involved more of their peers.

---

[26] As will be highlighted in the discussion, this sub-forum was created at the end of the course's first week.
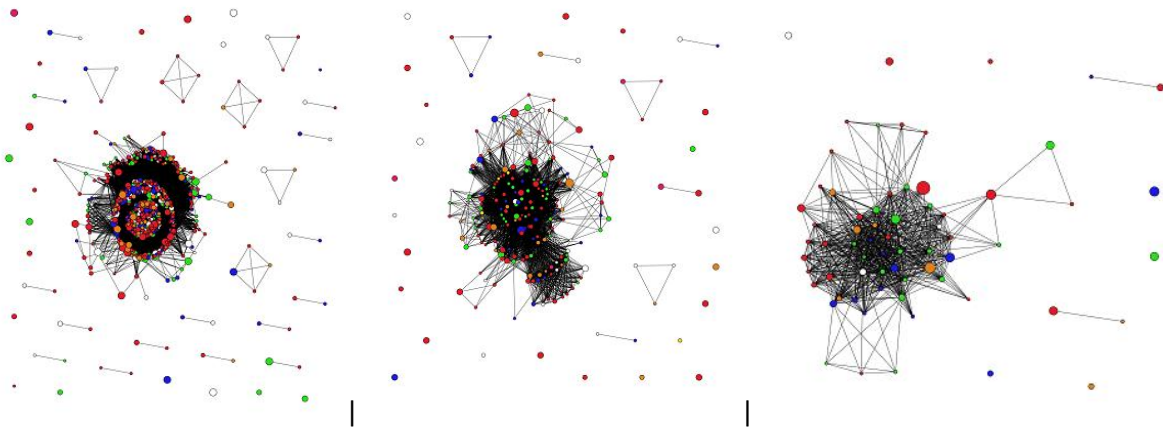
Figure 11: Thread participation in the Cases sub-forum during the first, third, and sixth weeks (left to right) of the course.

Summary

The thread-networks for weekly sub-forum activity reveal a number of insights about how discussions evolved throughout the course. While some sub-forums maintained dense communication centres as the course progressed (or saw dense communication at the beginning and end of the course), others started off with these high levels of discussion diversity and quickly fragmented into more topical, focused, and exclusive groups. The participants appeared to engage in crowd-like behaviour without forming long-lasting, collaborative communities (Stahl, et al., 2006) – a phenomenon to be explored further in the discussion. Additional trends emerged as alternative network topologies were explored.

Key structural trends observed via post-comment networks

Analysis of the post-comment networks enabled deeper insight into how individual participants interacted with one another than the thread networks. The post-comment networks for each sub-forum over the course's entire 7-week period offer the most insights into how participants interacted with one another in their real-world learning pursuits.

In the figures that follow, there appear to be nodes that are completely disconnected from others. These individuals have neither commented on another post, nor have had at least one of their posts commented on by their peers. Moreover, node sizes vary -- as described in the methodology, larger nodes correspond to forum participants that have higher levels of activity. Variances in node sizes imply forum participation volume equality across learners (Nielsen, et al., 2010).

Readings sub-forum

Figure 12 shows the post-comment relationships in the Readings sub-forum. The network contains tree-branch like structures, chains, and a few simple polygonal shapes. These structures are characteristic of simple information exchanges (Vaquero & Cebrian, 2013) where most users tend to directly interact with only a small number of others. Noticeably absent is a strong, dense discussion "core", implying that the sub-forum did not harbour sustained communication, but rather, one-off, demand-driven participation. This trend was confirmed by observing actual discussions during the course: in one case, a user posed a question about a very specific topic mentioned in the reading, generating responses from a number of other participants.

The average size of the nodes[27] was approximately 74 with a standard deviation of 17. This relatively small variance across node sizes implies somewhat equal participation volumes among users that discussed the readings.
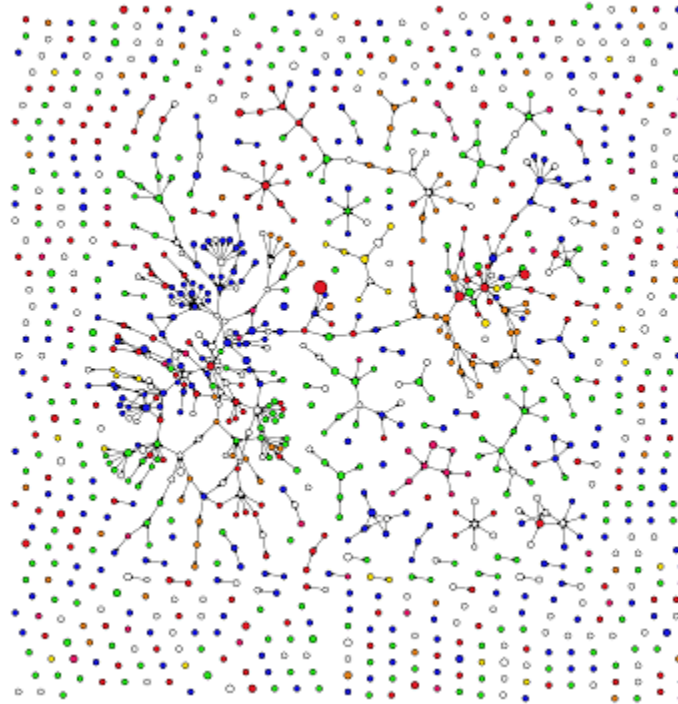

Figure 12: Post-comment participation in the Readings sub-forum.

Study Groups sub-forum

Many of the chain and tree-like structures that were present in the Readings network were also a part of study group formation. However, a key difference in this sub-forum was the existence of more prominent hub-and-spoke structures. Figure 13 shows some of the emergent trends. Many of these hub-and-spoke relationships, as verified by direct observations, told the following story: an individual created a thread and proposed a study group for a particular subset of participants (e.g. those in India or those that speak Spanish). A number of participants then converged on this thread and commented on the initial post to express interest. An assortativity analysis[28] on the tendency of individuals from the same continent to converse with one another revealed a Pearson r-value of .45, with $p < 0.001$ (Newman, 2002). This suggests that, with moderate strength, participants in this particular sub-forum tended to converse directly with other participants with which they share a geographic identity. Like the Reading sub-forum, participation volume across Study Groups sub-forum users had low variance, with an average node size and standard deviation of approximately 76 and 16, respectively.

---

[27] The methodology provides more details into how these node sizes were computed.
[28] This measures the tendency of nodes with similar attributes – such as continent of origin – to communicate with other nodes with the same or similar attributes.
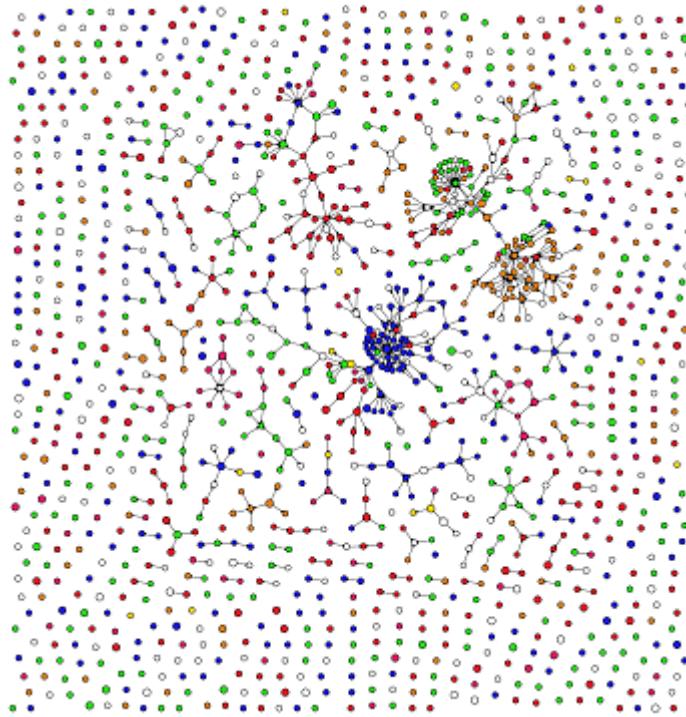
Figure 13: Post-comment participation in the Study Groups sub-forum.

Final Projects sub-forum

Figure 14 depicts post-comment relationships in this sub-forum. Similar to Readings, much participation in discussions about the final project was short-lived in nature, as evidence by the large number of isolated nodes as well as those nodes that form the end of a branch-like structure. Still, this network also appears to have a dense, well-defined centre of communication. As noted above, discussions in this sub-forum about questions regarding the final project write-up as well as how peer assessors should handle cases of plagiarism played a key role in creating discussion topics that joined a number of course participants. Once again, participation volume per user was relatively equal: the average node size was 82 and standard deviation approximately 20.
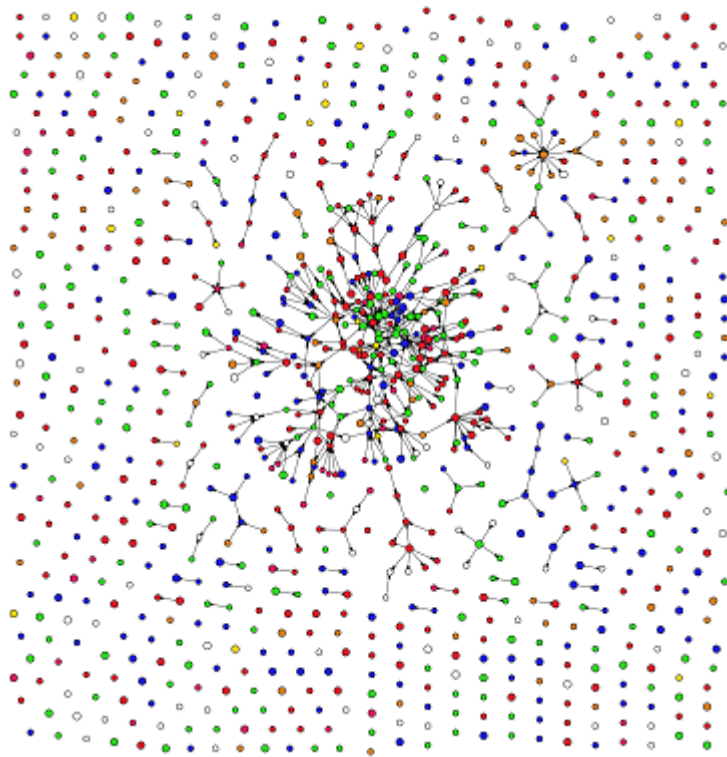
Figure 14: Post-comment participation in the Final Projects sub-forum.

Organizations Seeking Analyses sub-forum

Communication around the real-world issues put forth by organizations was less centralized than in the Final Projects sub-forum. Figure 15 shows a number of distinct groups with minimally overlapping participation that emerged in this setting. Similar to its parent sub-forum, participation levels in these discussions were relatively uniform: the standard deviation of nodes sizes was approximately 20 and the average node size was 83.
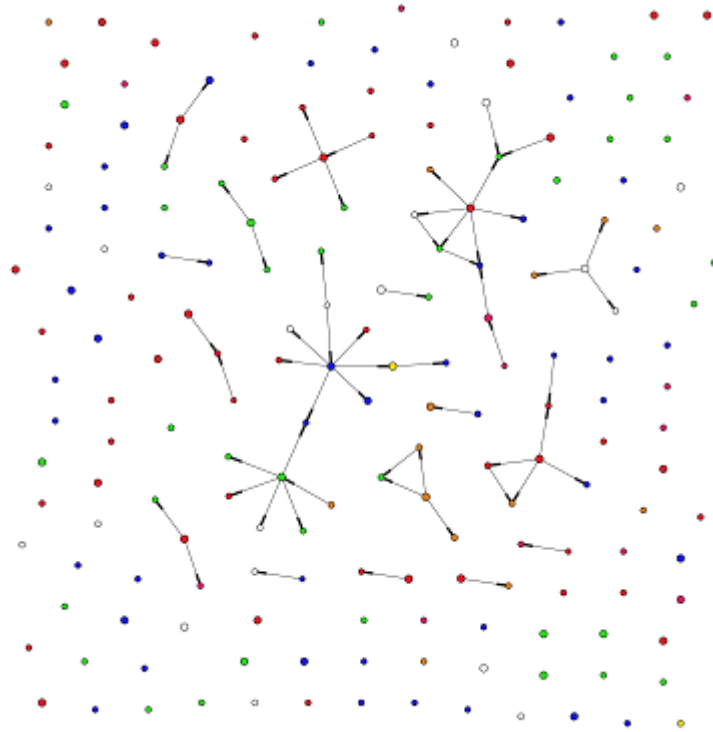
Figure 15: Post-comment participation in the Organizations Seeking Analysis sub-forum.

Cases sub-forum

Post-comment relationships in the Cases sub-forum were perhaps more diverse than in any other setting. Figure 16 illustrates the nature of these interactions. For one, communication in this context was very centralized: there were not as many isolated group discussions as in the previously-highlighted sub-forums, and many participants appeared to engage directly with many others. Moreover, there was much diversity in participation levels: the average node size was 156 with a standard deviation of 193. As evidenced by the large, isolated nodes, some individuals tended to participate heavily even without engaging directly with the posts of others, or having others engage directly with their posts.
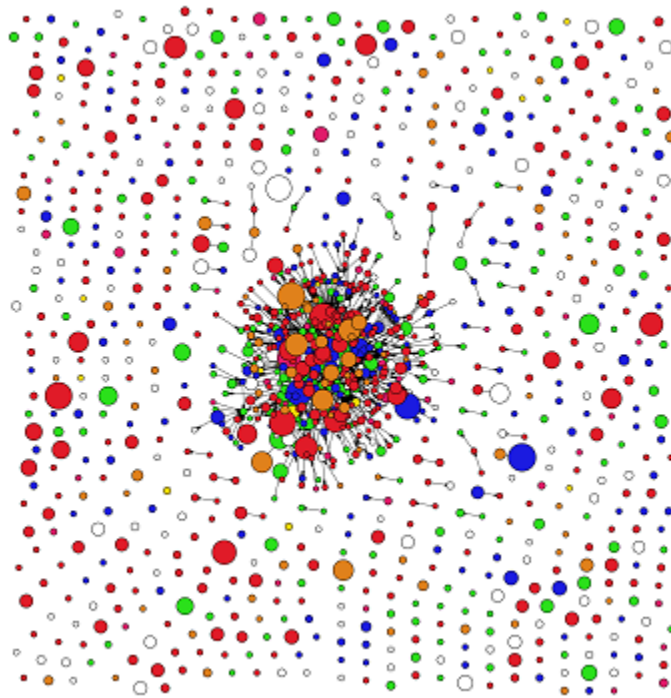
Figure 16: Post-comment participation in the Cases sub-forum.

The apparent centralization of discussions in this sub-forum naturally raises an important question: were there some people that were more *important* in these discussions than others? Without a full-scale content analysis, interviews, and post-hoc analysis of student outcomes it is impossible to answer this question completely. However, a coarse approximation of "importance" can be generated by computing the PageRank of each participant in the network. PageRank is an algorithm used to rank the relative importance of web pages (modelled as nodes) on the internet (modelled as a network). It assigns a value between 0 and 1 to each node (where the sum of all node values must be equal to 1) based on the directed relationships between nodes and has been used more generally to determine how central particular actors are in either communicating information to, or facilitating contact with, other actors in a network (Page, et al., 1999). In the context of post-comment relationships in this sub-forum, the PageRank of a participant indicates how central she is in communications with her peers.

In the Cases sub-forum, the highest PageRank of any given node was 0.07. To put this in perspective, given that there were 1,221 distinct participants in this sub-forum, if all participants assumed an equally important role in discussion, each would have a PageRank of approximately 0.0008. Therefore, the most "important" actor in this network had a PageRank nearly 100 times greater than the uniform case. Indeed, some students played a key role in facilitating discussions in response to questions with real-world relevance.

Summary

Post-comment networks revealed a variety of node relationships. Some discussions tended to be centralized while others were more distributed and fragmented. In some instances, such as the Cases sub-forum, participation volumes across nodes varied widely and some individuals played a highly important role in "transferring" information to and from others. These finds will be discussed further in

the forthcoming chapters.

Persistence of participation

With a broad understanding of how forum discussions in the MOOC evolved and how trends emerged across different network topologies, a fundamental question remains unanswered: how consistent was participation across these various sub-forums and weeks? The Jaccard Index is a simple measure of similarity between two sets that can help answer this question. For any sets A and B, $J(A, B) = \frac{|A \cap B|}{|A \cup B|}$, i.e., the Jaccard similarity index between A and B is equal to the ratio of items in both A and B to those items in A or B (Hamers, et al., 1989).

The Jaccard indices for pairwise forum participation reveal that participation was highly inconsistent. The largest overlap in the number of participants from any two sub-forums over the entire span of the course was .13, which occurred between the cases and final projects sub-forums. This means that these settings shared only 13% of the participants that engaged in either sub-forum. The remaining overlaps were much lower, dropping to as low as 1.4% similarity between the technical feedback and organizations seeking analyses settings. On a related note, the largest similarity between any two weekly networks over all sub-forums was .16 between weeks 5 and 6, and the lowest was .002 between weeks 1 and 7. Weekly overlaps within sub-forums demonstrated higher participation consistency in some instances: the participation overlap between weeks 4 and 6 and 5 and 6 in the cases sub-forum was .26 and .25, respectively. The numbers reveal that most participation in the forums was not due to the same few individuals in the course. It is interesting to consider the implications of these metrics, in conjunction with the tendency towards group fragmentation and focused discussions as exhibited by the network visualizations, on the existence of crowds versus communities in this MOOC.

Forum participation and performance in the course

The final stage of analysis included superimposing information about the final marks of students on the trends underlying their forum participation.

Forum participation by final score

It is important to analyse forum activity with respect to course outcomes in order to enable the interpretation of this work alongside previous studies. Of particular interest in final score data is that just over 20% of forum participants were individuals that earned 90% or higher in the course, even though only 2% of course participants overall finished with these scores.

To further understand the comparison of forum users' final scores with those of their other course counterparts, I computed a one-way analysis of variance using the Kruskal-Wallis test (Kruskal & Wallis, 1952) to compare the mean final score of forum users with that of another subset of users in the course[29]. I chose to compare forum users with those that watched or downloaded at least one lecture primarily because comparing against the larger body of course registrants - nearly 87,000 individuals - would have yielded mean final scores that were severely skewed due to a lack of active participation. The results indicated a strongly statistically and practically significant difference between the mean final scores of forum participants (24.4%) and those that participated in lectures (5.2%) (p = virtually 0.0, H-score = 2191). Practical significance (e.g. the effect size) was computed using Cohen's d, yielding a value of 4.0 (d > .8 implies practical significance). These values suggest that there is a

---

[29] This test was chosen because final grade distributions appeared more chi-squared than Gaussian. Still, the results must be caveated as the samples being compared were not independent due to the nature of students' participation in the course.

statistically and practically significant relationship between a student's participation in the forums and her performance in the course, although it cannot be concluded that forum participation *caused* higher final marks.

Conditional distributions of users' attributes and scores

The earlier discussion about forum users' background attributes can help illuminate the relationships between their characteristics and final marks. For one, over 35% of those that earned 90% or higher in the course were based in Europe -- the second highest cohort was approximately 26% from North America, followed by 23% from Asia. Europeans also comprised the highest proportion of forum users already with at least a master's degree. Unsurprisingly, then, those already with at least a master's degree also comprised the largest cohort (nearly 45%) of users receiving at least 90% in the class. Those forum participants with a bachelor's degree comprised the largest group of participants that earned 50% or less in the course.

Assortativity by final score

Previous literature has investigated the tendency of high-performing students to interact with other high-performing students in online learning settings (Vaquero & Cebrian, 2013). To measure the existence of similar trends in this MOOC's discussion forums, I computed the assortativity of each post-comment network with respect to participants' final marks. At best, there were weak correlations between participants' final scores and the final scores of the participants they tended to interact with. The highest final-score assortativity, or Pearson r-value, was .23 in the final projects forum. The lowest was .07 in the sub-forum dedicated to discussions about the video lectures. Table 7 shows these assortativity values for each sub-forum with corresponding 2-sided p-values. Note that due to the large sample sizes, all but one of these correlations are statistically significant at the $p < 0.05$ level. In the context of Cohen's effect sizes for social science research (Cohen, 1992), most of these correlations indicated, at best, relationships between communication and final scores of small to medium practical significance.

It is important to note that a lack of assortativity does not imply disassortativity, or a tendency for certain grade earners to interact with people from other grade-earning cohorts. Still, relatively weak assortativity values implies there is no strong indication that students who achieved final marks similar to others also favoured communication with those individuals.

Sub-forum	Pearson r-value (assortativity)	P-value	Cohen's effect size
Final Project	.23	p < 0.001	small to medium
Questions for Professor	.16	p < 0.05	small to medium
Readings	.12	p < 0.001	small to medium
Technical Feedback	.17	p < 0.05	small to medium
Lectures	.07	p < 0.1	Small
Cases	.08	p < 0.01	Small
Course Material Feedback	.22	p < 0.001	small to medium
Organizations Seeking Analyses	.22	p < 0.005	small to medium
Study Groups	.14	p < 0.001	small to medium

Table 7: Assortativity values, p-values, and effect sizes by learners' final marks, according to post-comment interactions in different sub-forums.

Summary

The results presented in this chapter offer one set of insights into how learners in this MOOC interacted through the discussion forums. The next chapter reflects on these results in light of the study's research questions.

Discussion

Forum participants are well-educated, primarily from the western world, and want to gain skills that will help them professionally

The results revealed that most forum participants (over 80%) had at least a Bachelor's degree and that 2 out of every 3 participants were from the western world. These trends are also characteristic of non-forum participants in the course and confirm findings from early literature: MOOCs, so far, have proven to be tools that promote continuing education among Americans and Europeans that already have substantial previous educational experiences (Breslow, et al., 2013; Kizilcec, et al., 2013). As MOOCs proliferate, if they are to fulfill the hopes of their advocates by widening participation in higher education, it will be important for them to accommodate and serve a truly global, heterogeneous population.

Overwhelmingly, "Professional Development" was the most popular reason for taking the course: over 93% of forum participants indicated this as their primary motivation, while most others (6%) were motivated by curiosity about the course or the pursuit of enjoyment. These statistics are very different than those reported for earlier MOOCs, such as Bioelectricity offered on Coursera through Duke University in 2012, which had a higher proportion of students primarily seeking enjoyment (Belanger & Thornton, 2013). Therefore, there is no single reason why people take MOOCs, nor why they participate in discussions within these courses. Their motivations are largely determined by the type of course and their own unique backgrounds and experiences.

Knowing forum participants' motivations for taking the course can also help learning technologists and educators develop tools that are better-equipped to accommodate them than what currently exists. Forum technologies have hardly evolved in the past 15 years, and in some cases, their limited functionality has failed to accommodate the diverse needs of students (Thomas, 2002). Indeed, as observations revealed, the fact that many forum participants in the MOOC tended to move away from the discussion forums to alternative technologies such as Facebook, Skype, and Email suggests a failure of the current tools in meeting the needs of a diverse student body. There are thus new opportunities to re-imagine how these online discussions can be facilitated within course environments themselves.

These trends, of course, are observations from one particular MOOC and are subject to the nuances of a particular pedagogical approach and set of learning objectives. Still, considering how forums may leverage, or perhaps even integrate with the technologies that MOOC participants use in their everyday lives is fundamental to understanding how discussions evolve in massive, open online classrooms. Indeed, the question of how "non-formal" or "informal" learning occurs via a vast array of online technologies is being investigated even beyond the higher education space (boyd, 2008; Davies & Good, 2009).

Students use discussion forums to form study groups and communicate about real-world topics

Trends in the most-viewed and most-posted-to threads and sub-forums revealed that students were particularly active in discussing real-world topics – which was in line with the course's focus on real-world applications of business practices. The Final Projects sub-forum was the most-viewed of all sub-forums, generating over 40,000 views, or 64% more than the second-ranked Study Groups setting. The weekly case discussions, particularly the first week's focus on Google and its market strategy, accounted for three of the top five most-viewed sub-forums. The Cases sub-forum varied the most in the extent to which each individual participated in discussions. In some cases, participants posted 500+ word analyses in response to one or more questions posed by the instructor. Moreover, as exemplified through the PageRank metric, some individuals in the Cases setting played a central role in numerous discussions that were conducted among other participants -- i.e., in some sense, they were "information hubs" through which discussions evolved. Long posts and centrality in the communication network were atypical, but they both indicate how some students played a key role in communication.

Students that participated in discussion forums were also driven by a desire to connect with their peers. The top 5 threads with the most posts were all devoted, in some way, to forming groups among participants or learning more about their motivations for taking the course. Social network analysis and informal observations confirmed these trends. Forum participation peaked in the first week, and much of the activity took part in the "Study Groups" sub-forum. As the assortativity analysis from the results section revealed, learners tended to seek out others with similar geographic backgrounds to form and sustain study groups.

A commonality across the Cases, Final Projects, and Study Groups sub-forums was that they seemed to encourage broad, open-ended discussions. The Cases forum, for example, posed a series of questions for students to answer publicly after reading each case study. Even though online learning has been criticized for years because of its alleged promotion of "mechanized" pedagogy (Nobel, 1998), this MOOC demonstrated how the discussion forums, and in particular, posing questions that require students to apply their learning to the real world can promote engagement and peer interaction – at least for those that choose to participate. Indeed, earlier literature in online learning has highlighted the importance of peer interactions via forums and chat-rooms to learning processes (Ryle & Cumming, 2007). Moreover, literature in problem-based learning (Saver & Duffy, 1995) has emphasized that the opportunity to apply one's learning to address challenges that capture real-world contexts can lead to greater cognitive engagement[30] in some instances (Rotgans & Schmidt, 2011). An important area of future investigation will include if, and how, real-world learning and problem solving via MOOC discussion forums can impact student engagement.

Because of the nature of the course, it is not surprising that forum participants appeared to engage with each other mostly around real-world issues. It is also not surprising that students who posted in the discussion forums were also proactive in forming study groups, since both are examples of social behaviour. Rather, this work helps cultivate a better understanding of how engagement differed at a more granular level of discussion and how discussion forums may be used to promote learner interactions in future courses. It will be important to validate the findings of this research against more robust content analysis of discussions within the identified threads. A deeper analysis will perhaps include a typological characterization of those that actually participated in popular discussions in order to better understand how they interacted with others (Nielsen, et al., 2010). This approach may strengthen current understanding of how different engagement patterns emerge around different discussion topics.

---

[30] Reported in this context through a validated self-assessment.

Alongside opportunities for future research, the findings of this study also encourage new efforts on the parts of software engineers and data scientists to develop real-time analytics tools for instructors. Recent reports (Bienkowski, et al., 2012) have pointed to the importance of capturing learner analytics – or, in the spirit of this study, data on how students interact with one another and other parts of their environments as they learned – to improve educational approaches and outcomes. Enabling course staff to identify relevant learner analytics as a course progresses could positively impact teaching practices during the course. For example, if instructors can identify forum participants with high "PageRank" - those that play a central role in facilitating communication with other students – it may help spread key learning objectives and messages more efficiently and effectively to other learners. While these tools will offer broad insights into the differing roles of students, it will be up to educators to determine how best to use this information to shape future pedagogical practices.

MOOC discussion forums harbour crowds, not communities of learners

Perhaps the largest benefit of employing social network analysis in visualizing communication in the forums was its elucidation of how different actors (learners) in different networks (sub-forums) exchanged information with one another. On the whole, the network images illustrated diminishing forum participation as the course progressed. As the weekly thread-networks demonstrated, groups that formed around particular discussion topics were generally short-lived and even those sub-forums that started off with highly centralized communication and diverse user activity quickly became fragmented and distributed. Even the Cases and Final Projects sub-forums, which maintained strong communication cores and had learners that participated in many different discussions, underwent periods of fragmentation and diminishing participation during certain parts of the course.

In some cases like in the Readings sub-forum, the post-comment networks illustrated a tendency for users to use the forums primarily to access information "on-demand" - i.e., to engage in interactions exclusively to either initiate or respond to dialogue among a select group of participants. In many instances, users would only comment on the post of a single other user, yielding "tree-branch", chain-like node and edge network structures. Most students did not interact with a large number of other students directly, but instead, focused on the types and topics of communication they wished to engage in and committed themselves for a short amount of time to those.

These insights reveal a fundamental characteristic of MOOC participants that engage in real-world discussions: they come together and disperse as crowds, not communities, of learners. As discussed in the literature review, Lave and Wenger have suggested the formation of "Communities of Practice" in situated learning settings, where community members are integrated from their early stages of "legitimate peripheral participation" to high-volume participation as they drive collaboration among other members (Lave, 1991). In the context of the MOOC's discussion forums, however, nearly the opposite was true: students started off with high-volume participation in online discussions and over time, decreased in their apparent commitment to these conversations.

Another trend observed consistently across the post-comment networks was the large number of nodes - particularly in the Cases sub-forum - that were very active in the forums but were completely disconnected from communication clusters. These individuals created threads and posted many times, but never once commented on the posts of their peers or received any comments on their own posts. It is important to note that these nodes did not exist to the same degree in the thread-networks, meaning they were not completely isolated in conversations. Moreover, it is possible that they "responded" to the posts of their peers by simply creating another post instead of adding a comment explicitly, and vice versa. Still, the existence of these nodes demonstrates that many participants were perfectly content with contributing their thoughts on various discussion topics without directly or explicitly engaging with others. They cooperated by virtue of the fact that they participated with other learners to engage in

mutual discussion (Panitz, 1996), but they likely did not collaborate with their peers to iteratively create knowledge or construct shared meaning through conversations.

The low proportion of students participating in the forums from week-to-week further illustrates the ephemeral nature of the crowds that formed in the course. The fact that only 0.2% of participants that used the forums in the beginning of the course also participated during the peer-review week - even though all forum users constituted nearly 5% of total course registrants – illuminates inconsistent discussion forum use.

While it may be appropriate to say that forum users tended to form crowds and not communities in the discussion forums themselves, it is unclear what happened on platforms outside of the MOOC. As noted in the results, observations revealed that many left the Coursera-based discussion forums to continue communications through other technologies, and in some cases, in-person. Indeed, in the case of some of the real-world final projects where organizations solicited the help of students, some students continued their conversations with others on Skype and email even after the course was over. Interviews with students will reveal the extent to which these groups built learning communities elsewhere.

It can be argued that much of the forum participation inconsistency on the MOOC platform is due to the fact that participation in the course itself is difficult to define. Since participants take MOOCs of their own volition, they are free to engage or disengage as they see fit. This confirms an important point: MOOCs are not like their more formal counterparts in online education, and therefore, they must not be considered in the same light or evaluated along the same dimensions. Particularly in the context of forum discussions, it will be important to develop new metrics for evaluating the contributions and achievements of learners in an individual and collective sense.

Higher-performing students use the discussion forums, but they do not only tend to interact with other higher-performing students

From the evidence it is clear that learners in this MOOC tended to form crowds, not communities, in the discussion forums. This behaviour not only sheds light on the uniqueness of MOOCs as educational settings but also calls for a re-imagination of how they may be used to promote learning within a global student body. Still, while many attributes of MOOCs break from precedents in formal education, most of these courses still assign final scores to participants so they may know how they fared overall. In the case of this MOOC, students were evaluated based on weekly quizzes (which they could resubmit any number of times for re-evaluation) and their final project analyses.

As mentioned earlier, this MOOC like many others had a high attrition rate: only 2,500 of the 90,000+ registrants received a nonzero final score. However, it is important to acknowledge that "attrition", or "dropping out", are inherently ambiguous in these settings. Since MOOCs are low-commitment and since many are yet to offer formal accreditation, they have so far primarily attracted individuals that are interested in continuing to learn and gain skills instead of those seeking formal recognition for their work (Kizilcec, et al., 2013). There is no comprehensive or standardized metric, then, to assess whether each student has learned what she initially took the course to learn.

Therefore, students' final scores in these courses only partially explain enrolment, participation, and achievement. Still, my analysis of the final scores of forum participants in particular offered a deeper understanding of how and why students conversed with one another. The fact that the forums had a statistically and practically significant proportion of high grade earners in comparison to the rest of the course - as exhibited by the fact that 20% of forum participants received at least a final score of 90% - indicates that forum participation revealed something about students'
performance. Observations confirmed that many students used the forums to ask specific questions about assignment due dates and requirements, for example. A recent paper by researchers at Harvard

and MIT on data from an edX MOOC revealed a similar tendency of course certificate earners (generally high-scoring individuals) to turn to the discussion forums for help in solving homework problems (Breslow, et al., 2013). For the business MOOC, the converse statistic was also compelling: 44% of those individuals that received at least a 90% in the course also participated in the forums. Indeed, many forum participants across MOOCs have proven to be a high-performing subset of the rest of the course.

It is important to consider *who* these high score-earning forum participants were. The results indicated that Europe-based forum participants were both the highest grade-earners and held the highest proportion of master's/professional degrees - even though they comprised only 25% of total forum participants. This confirms trends observed in earlier MOOCs: well-educated people from the western world tend to complete MOOCs and do well in them. However, there are additional trends to note. North Americans comprised 32% of forum participants but only accounted for 26% of those forum participants receiving at least a 90%, while Asians accounted for 24% of forum participants but 23% of 90%+ earners. This suggests that a higher proportion of forum participants from Asia than North America received high marks. While this data is certainly not conclusive, it reveals one possibility of MOOCs: their ability to enable high-performing students from the non-western world to discuss real-world topics with their counterparts elsewhere.

High performers constituted a substantial proportion of forum participants, but certainly not a majority; it is perhaps more revelatory that over 73% of forum participants received below a 50%, and these participants accounted for nearly 60% of all forum comments and posts. This suggests that most discussion forum participants are more interested in connecting with others to talk about issues with real-world significance and implications than they are in being formally recognized for their work. Previous investigations of online learning communities have indeed revealed that those that are most active in forum discussions are not necessarily the highest performing students (Davies & Graff, 2005). By many conventional standards, it is bizarre that such a large amount of forum activity was attributed to those that "failed" the course. However, the data reveals that final scores are not a sufficient measure of student ability or performance as they fail to reflect why so many people do so poorly but are nevertheless so active in the forums. Many forum participants are likely intrinsically motivated to participate in discussions, which reflects trends in crowdsourcing where some actors are driven by a desire to give back to a problem or set of ideas greater than themselves (Budhathoki & Haythornthwaite, 2012).

From these results, we can conclude that participation in the forums was dominated by people who received very high or very low final marks. But in most cases, people with diverse final-scores tended to interact with one another. The assortativity analysis on post-comment networks revealed that there was at best a medium, and in general, relatively small correlation and corresponding effect size in how similarly-scoring participants tended to communicate with one another. This finding is contrary to earlier literature that indicated high-performing students tend to interact with other high-performing students early on in online courses (Vaquero & Cebrian, 2013). MOOCs, then, have the potential to not only enable communication between learners from different parts of the world, but also learners with different motivations and skill-levels. There is much literature that has already confirmed this thought in the context of other learning settings: Social Interdependence theorists have already begun to formalize how educators can encourage group formation and dialogue among a body of learners (Johnson & Johnson, 2009). Given this grounding and the previously noted results, it is appropriate to view the ability of MOOCs to bring together geographically, culturally, and intellectually diverse groups of people as one of their most valuable contributions to education.

Harnessing the power of diverse conversationalists in online, real-world learning settings

The forum participants in this MOOC formed crowds, not communities, of conversationalists that engaged in a variety of discussions. Visualizations of their communications revealed that many individuals - especially those that did not receive a passing mark in the course - participated in many high-volume discussions. These participants appeared to seek connections with their peers but were not necessarily motivated by formal assignments or lectures. Instead, many engaged because they sought experiences and insights that would help them in their offline/professional lives.

Participants used the forums to talk about issues ranging from forming study groups with one another to challenging each other about how best to provide companies like Google with business strategy advice. Some students even connected with existing organizations to apply the course's tools and frameworks to real-world business strategy challenges.

Without stringent academic requirements, MOOCs resemble non-formal learning environments in that many students decide to participate simply to further their knowledge (Colley, et al., 2002). This research, as well as early literature on MOOCs, confirm that students in MOOCs are lifelong learners primarily motivated by a desire to continue their education. As this course's forum discussions revealed, a majority are not in pursuit of formal credentials or marks that signify formal academic achievement. An objective of future research in MOOCs should be to understand how more suitable methods of evaluation – instead of final marks based solely on quiz and assignment participation – can be employed in these non-formal contexts. These methods of evaluation will need to differ according to the topic of the MOOC and the nature of the pedagogical practices employed. As this study reveals, how learners interact through the forum discussions, especially around real-world issues, may offer one dimension along which alternative assessment strategies could be explored.

The research agenda for MOOCs is growing as new questions arise about their pedagogical value. Fine-grained data on learners' interactions with online course platforms – and each other – offer hope for investigating and understanding how participation differs according to demographic and environmental attributes. Moreover, some MOOCs that incorporate real-world learning topics bring together young adults with previous work experiences, seeking skills development and professional advancement. It is yet to be seen if the results of this study carry over to discussion settings in other real-world learning MOOCs – or other MOOCs more broadly – but these early insights indicate that understanding interactions between students may enable educators to design more engaging and effective learning opportunities in the future.

Conclusion

This research has started to uncover which types of students in MOOCs emphasizing real-world topics tend to interact with one another, and how. The study comes at a time when MOOC research is in its infancy, with recent papers focusing on exploring broad trends in learner participation and achievement – without much focus, yet, on communication trends. Sociocultural theories of education have pointed to peer-communication - particularly discussions across various cultural and geographic boundaries - as important to unlocking the learner's potential (Vygotsky, 1978; Leont'ev, 1978). In the context of online interactions across a variety of professional and educational settings, both theorists and practitioners have tried to distinguish between cooperation and collaboration (Stahl, et al., 2006), and correspondingly, crowds of actors that engage in less-committal interactions versus communities of people that come together to construct shared meanings and insights (Haythornthwaite, 2009).

While MOOCs combine characteristics of online classrooms and large-scale crowdsourcing platforms, early reports have not yet explored if their participants tend to assemble into crowds or communities. They also have not revealed *why* learners appear to use the discussion forums to connect

with their peers from around the world. Instead, research to date has focused primarily on offering a foundational look at statistics about course participants, their backgrounds, and overall performance. However, it is important to note that the discussion forums – which bring together thousands of people from around the world in a semi-synchronous setting – are a key component of what distinguishes MOOCs from previous incarnations of online learning. Therefore, investigating how interactions occur in the forums may offer new insights into the pedagogical value and potential of MOOCs.

Using surveys, users' digital traces, and observational notes from a business strategy MOOC that incorporated real-world business issues, I sought insights into how communication in the course evolved and the different roles participants played in these communications. The dataset was robust: nearly 87,000 people were enrolled at the end of the course and the digital traces included information on their nuanced behaviours both within and beyond the discussion forums. In order to better understanding how they used the forums, I employed social network analysis (SNA) to define and visualize communication networks. By looking at networks as they evolved over different discussion topics and time windows, I was able to learn not only about the backgrounds of forum participants, but also when and how they interacted with one another, and what forum participation revealed about their final marks in the course. This hybrid methodology helped yield both qualitative and quantitative insights into learner participation and communication in this setting.

Overall, most forum participants were well-educated, from Europe or North America, and were taking the course to gain professional skills. The most popular discussions tended to focus on real-world topics such as case analyses of large companies like Google. Students also used the forums to connect with one another to form study groups, which were often delineated *de-facto* according to geography or language. The groups that formed in the forums resembled crowds and not communities: some students engaged in rich discussions around topics of shared significance, but largely, forum use entailed short posts where students tended to access information "on-demand". Many students also grew more focused in their discussion participation after a tendency to discuss many different topics in the early parts of the course.

A sizeable proportion of discussion participants was comprised of high-performing students, although most others received what would have been considered as "failing" marks. This is of particular significance, as it reveals that the vast majority of forum participants were not driven by formal certification or acknowledgement of course participation -- instead, they appeared to engage because they were interested in the discussion topics and wanted to gain skills that would help them in their lives outside of the course. This trend suggests how MOOCs may evolve. Perhaps their true value lies not in how they deliver educational content to a dispersed audience, but in how they enable and encourage discussions between members of this audience. Understanding the role these discussions can play in students' learning - and how they may have an impact on the realities of participants - will be important to the future of MOOCs.

There is much further work to be done to gain fundamental insights into how MOOC participants use the discussion forums. First, the SNA techniques used in this study could be refined in a number of ways. Currently, the network visualizations do not capture information about the content of discussions or the details of how individuals communicated with one another. In the future, it would be interesting to manipulate additional network attributes - such as edge weights – to gain insights into the "importance" of certain communication channels. Researchers should also consider alternative network topologies - for example, where nodes represent discussion threads instead of individual participants and edges indicate some measure of similarity between threads. These representations could supplement larger scale content-analysis and participant interviews for a more robust understanding of the nature of participant interactions.

As highlighted in the discussion section, the MOOC platform's forums are incomplete depictions of how people formed groups and interacted with one another in the course. In the future, it will be important to augment MOOC-specific datasets with data from discussion groups created on other online platforms. The course also revealed that many participants preferred to connect with one another using social media; it may be possible, using publicly exposed programming functionality, to parse the data streams from these various media outlets to build a more complete picture of students' communications with one another. A more complete dataset of student interactions and how they persist outside of the MOOC setting itself will enable instructors and policymakers to design more robust learning experiences moving forward.

There are a number of research projects that could be generated simply by focusing on the real-world learning aspects of the final project. Analysing the content of students' final project submissions and case discussions, for example, could help reveal the learner traits that are most indicative of high performance. More specifically, one could explore how students engaged with real-world organizations to address live business strategy challenges. Investigating this subset of forum users could offer insights into the role a given student's previous life experiences and backgrounds play in her contributions to these settings – and how MOOC technologies themselves can be used to mediate more effective interactions, similar to previous explorations in professional settings (Kim, et al., 2008). Moreover, it could help educationalists learn more about the links between learning, real-world relevance of content, and communication. Given the diversity of MOOC participants to date, it is of particular importance to learn how future courses may better embrace this diversity and create environments that enable these learners to thrive.

The current discussions in higher education circles and popular news outlets about MOOCs is only the beginning. These courses may or may not offer sustainable opportunities for education in the developing world. Moreover, their pedagogical value is still being realized and how they will integrate themselves with existing paradigms of higher education is still unclear. But these open issues should not cloud a deeper investigation into how MOOCs can promote learning across geographic divides. It is time to shift the MOOC debate from questions about how they compare to campus-based education to how their unique contributions – namely, their promotion of global-scale conversations – can be used to help improve learning. Understanding how lifelong learners from around the world connect with their peers to ask questions and promote dialogue, often about topics that relate to their lives beyond the MOOC, is an important first step in exploring how these courses can meet the high expectations set forth to date. Perhaps this will be the value add of MOOCs - bringing people with shared intellectual interests together to engage directly with one another around real-world issues.

References

Abelson, H., 2008. The Creation of OpenCourseWare at MIT. *Journal of Science, Education, and Technology,* 17(2), pp. 164-174.

Baldwin, T., Bedell, M. D. & Johnson, J., 1997. The Social Fabric of a Team-Based M.B.A. Program: Network Effects on Student Satisfaction and Performance. *The Academy of Management Journal,* 40(6), pp. 1369-1397.

Barabási, A., 2009. Scale-Free Networks: A Decade and Beyond. *Science,* 325(5939), pp. 412-413.

Barab, S., 2004. Design-Based Research: Putting a Stake in the Ground. *The Journal of the Learning Sciences,* 13(1), pp. 1-14.

Bassey, M., 1999. *Case Study Research in Educational Settings.* Philadelphia: Open University Press.

Belanger, Y. & Thornton, J., 2013. *Biolectricity: A Quantitative Approach,* Durham, NC: Duke Center for Instructional Technology.

Bennet, S. & Oliver, M., 2011. Talking back to theory: the missed opportunities in learning technology research. *Research in Learning Technology,* 19(3), pp. 179-189.

Bienkowski, M., Feng, M. & Means, B., 2012. *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief,* Washington, D.C.: U.S. Department of Education, Office of Educational Technology.

Bogdan, R. & Biklen, S. K., 1982. *Qualitative Research for Education.* Boston: Pearson Education.

Bork, A., 1995. Distance Learning and Interaction: Toward a Virtual Learning Institution. *Journal of Science Education and Technology,* 4(3), pp. 227-244.

boyd, d. & Crawford, K., 2012. Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society,* 15(5), pp. 662-679.

Breslow, L. et al., 2013. Studying Learning in the Worldwide Classroom: Research into edX's First MOOC. *Research & Practice in Assessment,* Volume 8, pp. 13-25.

British Educational Research Association, 2011. *Ethical Guidelines for Educational Research,* London: s.n.

Brown, J. S., Collins, A. & Duguid, P., 1989. Situated Cognition and the Culture of Learning. *Educational Researcher,* 18(1), pp. 32-42.

Buchanan, M. & Calderalli, G., 2010. A networked world. *Physics World,* 23(2), pp. 22-24.

Budhathoki, N. & Haythornthwaite, C., 2012. *Motivation for Open Collaboration: Crowd and Community Models and the case of OpenStreetMap,* s.l.: Unpublished draft.

Cho, H., Gay, G., Davidson, B. & Ingraffea, A., 2007. Social networks, communication styles, and learning performance in a CSCL community. *Computers & Education,* pp. 309-329.

Cohen, J., 1992. Statistical Power Analysis. *Current Directions in Pyschological Science,* 1(3), pp. 98-101.

Colley, H., Hodkinson, P. & Malcolm, J., 2002. *Non-formal learning: mapping the conceptual terrain,* Leeds: University of Leeds Lifelong Learning Institute.

Daniel, J., 2012. *Making Sense of MOOCs: Musings in a Maze of Myth, Paradox, and Possibility,* Seoul, South Korea: Academic Partnerships.

Davies, J. & Graff, M., 2005. Performance in e-learning: online participation and student grades. *British Journal of Educational Technology,* Volume 36, pp. 657-663.

Deboer, J., Breslow, L., Stump, G. & Seaton, D., 2013. Diversity in MOOC Students' Backgrounds and Behaviours in Relationship to Performance in 6.002x. *In press..*

Easley, D. & Kleinberg, J., 2010. *Networks, Crowds, Markets: Reasoning about a Highly Connected World.* Cambridge: Cambridge University Press.

Emerson, R., Fretz, R. & Shaw, L., 2001. Participant Observation and Fieldnotes. In: P. Atkinson, et al. eds. *Handbook of Ethnography.* London: Sage, pp. 352-368.

Engestrom, Y., 2008. Expansive Learning: Toward an activity-theoretical reconceptualization. In: K. Illeris, ed. *Contemporary Theories of Learning: Learning Theorists in Their Own Words.* Hoboken: Routledge, pp. 53-73.

Fenn, J., Raskino, M. & Gammage, B., 2009. *Gartner's Hype Cycle Special Report for 2009,* s.l.: Gartner, Inc..

Hamers, L. et al., 1989. Similarity measures in scientometric research: The Jaccard index versus Salton's cosine formula. *Information Processing & Management,* 25(3), pp. 315-318.

Haythornthwaite, C., 1996. Social Network Analysis: An Approach and Technique for the Study of Information Exchange. *Library and Information Science Research,* 18(4), pp. 323-342.

Haythornthwaite, C., 2002. Building Social Networks Via Computer Networks: Creating and Sustaining Distributed Learning Communities. In: K. A. Renninger & W. Shumar, eds. *Building Virtual Communities: Learning and Change in Cyberspace.* Cambridge: Cambridge University Press, pp. 159-190.

Haythornthwaite, C., 2009. *Crowds and Communities: Light and Heavyweight Models of Peer Production.* Big Island, Hawaii, Hawaii International Conference on System Sciences.

Haythornthwaite, C., 2009. *Online Knowledge Crowds and Communities.* Reno, International Conference on Knowledge Communities.

Hedges, L. & Olkin, I., 1987. Statistical Methods for Meta-Analysis. *Journal of the American Statistical Association,* 82(397), pp. 350-351.

Herrington, T. & Herrington, J., 2006. *Authentic Learning Environments in Higher Education.* Hershey, PA: Idea Group Inc..

Huerta-Quintanilla, R., Canto-Lugo, E. & Viga-de Alva, D., 2013. Modeling Social Network Topologies in Elementary Schools. *PLOS One,* 8(2), pp. 1-9.

Hughes, G., 2010. Identity and belonging in social learning groups: the importance of distinguishing social, operational and knowledge-related identity congruence. *British Eduacational Research Journal,* 36(1), pp. 47-63.

Johnson, D. W. & Johnson, R. T., 2009. An Educational Psychological Success Story: Social Interdependence Theory and Cooperative Learning. *Educational Researcher,* 38(5), pp. 365-379.

Johnstone, S., 2005. Open Educational Resources Serve the World. *Educase*, January.

Kim, T., Chang, A., Holland, L. & Pentland, A., 2008. *Meeting Mediator: Enhancing Group Collaboration using Sociometric Feedback.* New York, NY, Proceedings of the 2008 ACM conference on Computer supported cooperative work.

Kizilcec, R., Piece, C. & Schneider, E., 2013. *Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses.* Leuven, Belgium, The 3rd Proceedings of the Learning Analytics & Knowledge Conference.

Kossinets, G. & Watts, D., 2006. Empirical Analysis of an Evolving Social Network. *Science,* Volume 311, pp. 88-90.

Krippendorff, K., 2013. *Content Analysis: An Introduction to its Methodology.* 3rd ed. London: Sage.

Kruskal, W. & Wallis, A., 1952. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association,* 47(260), pp. 583-621.

Larreamendy-Joerns, J. & Leinhardt, G., 2006. Going the Distance with Online Education. *Review of Educational Research,* 76(4), pp. 567-605.

Lave, J., 1991. Situated learning in communities of practice.. In: L. B. Resnick, J. M. Levine & S. D. Teasley, eds. *Perspectives on socially shared cognition.* Washington, D.C.: American Psychological Association, pp. 63-82.

Lave, J. & Wenger, E., 1991. *Situated learning: Legitimate peripheral participation.* Cambridge: Cambridge University Press.

Leont'ev, A., 1978. *Activity, Consciousness, and Personality.* s.l.:Prentice-Hall.

Limpert, E., Stahel, W. A. & Abbt, M., 2001. Log-normal Distributions across the Sciences: Keys and Clues. *BioScience,* 51(5), pp. 341-352.

Lincoln, Y., Lynham, S. & Guba, E., 2011. Paradigmatic Controversies, Contradictions, and Emerging Confluences, Revisited. In: N. Denzin & Y. Lincoln, eds. *The SAGE Handbook of Qualitative Research .* London: Sage, pp. 97-128.

Littlejohn, A., 2013. *Understanding Massive Open Online Courses,* Glasgow: Glasgow Caledonian University.

Markham, A. & Buchanan, E., 2012. *Ethical Decision-Making and Internet Research: Version 2.0,* s.l.: AOIR Ethics Committee.

McKinney, D., Dyck, J. L. & Luber, E. S., 2008. iTunes University and the classroom: Can podcasts replace Professors?. *Computers & Education,* Volume 52, pp. 617-623.

Moor, J. H., 1997. Towards a Theory of Privacy in the Information Age. *Computers and Society,* pp. 27-32.

Newman, M. J., 2002. Assortative Mixing in Networks. *Physical Review Letters,* 89(20), pp. 1-4.

Nielsen, W., Chan, E. & Jahng, N., 2010. Collaborative learning in an online course: A comparison of communication patterns in small and whole group activities. *The Journal of Distance Education,* 24(2), pp. 39-58.

Niles, H., 1922. Correlation, Causation and Wright's Theory of "Path Coefficients". *Genetics,* 7(3), pp. 258-273.

Nobel, D. F., 1998. Digital diploma mills: The automation of higher education. *Science as Culture,* 7(3), pp. 355-368.

Page, L., Brin, S., Motwani, R. & Winograd, T., 1999. *The PageRank Citation Ranking: Bringing Order to the Web,* Stanford, CA, USA: Stanford InfoLab.

Palonen, T. & Hakkarainen, K., 2000. Patterns of Interaction in Computer-Supported Learning: A Social Network Analysis. In: B. Fishman & S. O'Conner-Divelbiss, eds. *Fourth International Conference of the Learning Sciences.* Mahwah, NJ: Erlbaum, pp. 334-339.

Panitz, T., 1996. *Collaborative versus Cooperative Learning: A Comparison of the Two Concepts Which Will Help Us Understand the Underlying Nature of Interactive Learning.* [Online]
Available at: http://ericae.net/k12assess/colcoo.html
[Accessed 21 7 2013].

Preece, J., Nonnecke, B. & Andrews, D., 2004. The top five reasons for lurking: improving community experiences for everyone. *Computers in Human Behavior,* 20(2), pp. 201-223.

Psorakis, I., Roberts, S., Rezek, I. & Sheldon, B., 2012. Inferring social network structure in ecological systems from spatio-temporal data streams. *Journal of the Royal Society,* 9(76), pp. 3055-3066.

Punch, K. F., 2003. *Survey Research: The basics.* London: Sage.

Punch, K. F., 2009. *Introduction to Research Methods in Education.* London: Sage.

Roschelle, J. & Teasley, S., 1995. The Construction of Shared Knowledge in Collaborative Problem Solving. In: C. O'Malley, ed. *Computer Supported Collaborative Learning.* s.l.:Springer Berlin Heidelberg, pp. 69-97.

Rotgans, J. I. & Schmidt, H. G., 2011. Cognitive engagement in the problem-based learning classroom. *Adv Health Sci Educ Theory Pract,* 16(4), pp. 465-479.

Ryan, L., 2013. *MOOCs are on the Move: A Snapshot of the Rapid Growth of MOOCs,* s.l.: EFMD.

Ryle, A. & Cumming, K., 2007. Reflections on Engagement in Online Learning Communities. *International Journal of Pedagogies and Learning,* 3(3), pp. 35-46.

Saver, J. R. & Duffy, T. M., 1995. Problem Based Learning: An instructional model and its constructivist framework. *Educational Technology,* Volume 35, pp. 31-38.

Savery, J. & Duffy, T., 1995. Problem Based Learning: An instructional model and its constructivist framework. *Educational Technology,* pp. 31-38.

Seaton, D. et al., 2013. Who Does What in a Massive Open Online Course?. *In press..*

Siemens, G., 2005. Connectivism: A Learning for a Digital Age. *International Journal of Instructional Technology and Distance Learning,* 2(1), pp. 3-10.

Slingsby, A., Beecham, R. & Wood, J., 2012. *Visual analysis of social networks in space and time,* Newcastle, UK: Mobile Data Challenge 2012 (Nokia).

Stahl, G., Koschmann, T. & Suthers, D., 2006. Computer-supported collaborative learning: An historical perspective.. In: R. K. Sawyer, ed. *Cambridge handbook of the learning sciences.* Cambridge, UK: Cambridge University Press, pp. 409-426.

Thomas, M. J. W., 2002. Learning within incoherent structures: the space of online discussion forums. *Journal of Computer Assisted Learning,* Volume 18, pp. 351-366.

Vaquero, L. & Cebrian, M., 2013. The rich club phenomenon in the classroom. *Nature: Scientific Reports,* pp. 1-8.

Vygotsky, L., 1978. *Mind in Society.* Cambridge, Massachusetts: Harvard University Press.

Wasserman, S. & Faust, K., 1994. *Social Network Analysis.* Cambridge: Cambridge University Press.

Wiley, D., 2007. *On the Sustainability of Open Educational Resource Initiatives in Higher Education,* s.l.: OECD.

Wilson, J., Goodman, P. & Cronin, M., 2007. Group Learning. *The Academy of Management Review,* 32(4), pp. 1041-1059.

Yin, R. K., 2014. *Case Study Research: Design and Methods.* 5th ed. London: Sage.