What do cMOOC participants talk about in Social Media? A Topic Analysis of Discourse in a cMOOC

Citation for published version:
Joksimovic, S, Kovanovic, V, Jovanovic, J, Zouaq, A, Gasevic, D & Hatala, M 2015, What do cMOOC participants talk about in Social Media? A Topic Analysis of Discourse in a cMOOC. in *Proceedings of the 5th International Conference on Learning Analytics & Knowledge (LAK 2015).* ACM Press, New York, NY, USA, pp. 156-165. https://doi.org/10.1145/2723576.2723609

Digital Object Identifier (DOI):
10.1145/2723576.2723609

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Proceedings of the 5th International Conference on Learning Analytics & Knowledge (LAK 2015)

What do cMOOC participants talk about in Social Media?

A Topic Analysis of Discourse in a cMOOC

Srećko Joksimović
School of Interactive Arts and Technology
Simon Fraser University
Burnaby, Canada
sjoksimo@sfu.ca

Vitomir Kovanović
School of Informatics
The University of Edinburgh
Edinburgh, UK
v.kovanovic@ed.ac.uk

Jelena Jovanović
Department of Software Engineering
Faculty of Organizational Sciences
Belgrade, Serbia
jeljov@fon.rs

Amal Zouaq
Department of Mathematics and Computer Science
Royal Military College of Canada
Kingston, Canada
amal.zouaq@rmc.ca

Dragan Gašević
Schools of Education and Informatics
The University of Edinburgh
Edinburgh, UK
dgasevic@acm.org

Marek Hatala
School of Interactive Arts and Technology
Simon Fraser University
Burnaby, Canada
mhatala@sfu.ca

ABSTRACT

Creating meaning from a wide variety of available information and being able to choose what to learn are highly relevant skills for learning in a connectivist setting. In this work, various approaches have been utilized to gain insights into learning processes occurring within a network of learners and understand the factors that shape learners' interests and the topics to which learners devote a significant attention. This study combines different methods to develop a scalable analytic approach for a comprehensive analysis of learners' discourse in a connectivist massive open online course (cMOOC). By linking techniques for semantic annotation and graph analysis with a qualitative analysis of learner-generated discourse, we examined how social media platforms (blogs, Twitter, and Facebook) and course recommendations influence content creation and topics discussed within a cMOOC. Our findings indicate that learners tend to focus on several prominent topics that emerge very quickly in the course. They maintain that focus, with some exceptions, throughout the course, regardless of readings suggested by the instructor. Moreover, the topics discussed across different social media differ, which can likely be attributed to the affordances of different media. Finally, our results indicate a relatively low level of cohesion in the topics discussed which might be an indicator of a diversity of the conceptual coverage discussed by the course participants.

Categories and Subject Descriptors

J.1 [Administrative Data Processing] Education; K.3.1 [Computer Uses in Education] Distance learning

LAK '15, March 16 - 20, 2015, Poughkeepsie, NY, USA
Copyright 2015 ACM 978-1-4503-3417-4/15/03…$15.00
http://dx.doi.org/10.1145/2723576.2723609

General Terms

Human Factors, Algorithms

Keywords

Connectivism, Content analysis, SNA, cMOOC

1. INTRODUCTION

The initial development of Massive Open Online Courses (MOOCs) dates back to 2005, and coincides with the ideas of connectivism and networked learning [1]. While the first publicly available MOOC was the Connectivism and Connective Knowledge (CCK08) course in 2008, it was in 2011 when MOOCs started gaining significant attention [2]. Although MOOCs very quickly became an important component of the adult online education, there is presently an extensive debate about their role in higher education [3, 4]. The main concerns are related to the effective scaling-up of traditional courses and the ability of MOOCs and their underlying pedagogy to meet the needs of higher education [3].

Within the last several years, two prominent types of MOOCs evolved. The more centralized type of MOOCs – xMOOCs – are focused on content delivery to large audiences, where the learning process is teacher-centered, i.e., based on transferring knowledge from instructors to learners [5]. xMOOCs are usually delivered using a single platform (learning management system), where learners receive knowledge (most commonly in a video format), and further apply that knowledge in projects defined by the teacher [5]. On the other side of the spectrum, more distributed MOOCs emerged (cMOOCs). In cMOOCs, teachers' role is primarily focused on the early instructional design and facilitation. cMOOCs do not rely on any centralized platform but rather use various social media for sharing information and resources among learners. The main goal of learning in cMOOCs is knowledge building through connection and collaboration with peers [6]. Learners are co-creators of the content and there is no formal evaluation of the learning achievements.

The most commonly indicated issues and challenges related to MOOCs are low course completion rates, high degree of learner attrition, and the lack of a theoretical framework that would allow for better understanding of learning processes in networked learning [7]. In their analysis of the research proposals submitted to the MOOC Research Initiative[1] (MRI), [7] showed a promising upturn in addressing a wide variety of the challenges recognized to date. Majority of submissions proposed well-established frameworks in educational research and social sciences as a foundation for examining and understanding learner motivation, metacognitive skills, and other factors that shape learning and teaching in MOOCs.

However, our literature review indicates that most of the current studies on cMOOCs are based on quantitative methods and rather simple metrics (e.g., the frequency of facilitators' and learners' postings) [8, 9]. Without the capacity to explain practice and complexity of networked learning, existing approaches and research models do not allow for understanding of learning at scale [10]. To contribute to the current research practices in this area, our study proposes a combined use of automated content analysis and social network analysis (SNA) in order to provide a more effective approach to MOOC research. More precisely, the study reported in this paper suggests an analytic method that integrates quantitative (automated content analysis and SNA) and qualitative analysis of posts created within different social media platforms used in a cMOOC. Relying on tools for automated concepts extraction, as well as SNA tools and techniques, we were able to identify main groups of concepts emerging from learners' posts and to analyze how they evolve throughout the course. Further qualitative analysis enabled a more in-depth interpretation of our findings.

Having that cMOOCs often incorporate various technologies into the learning process, our first objective was to examine how different social media influence the discourse of course participants. The second objective was related to the role of course facilitators in a cMOOC. More precisely, our objective was to analyze how course readings, suggested by course facilitators, frame the topics being discussed among learners. Finally, we were interested in analyzing learners' discourse through a temporal dimension, that is, how topics discussed by students changed over time, when certain topics emerged and whether we can identify topics that sustained throughout the course.

2. THEORETICAL BACKGROUND AND RESEARCH QUESTIONS

2.1 Connectivism and cMOOCs

The theoretical foundation behind cMOOCs is connectivism [1, 11] and its principles of autonomy, diversity, openness and interactivity [12]. Connectivism is proposed as a novel theory of learning for "the digital age" [13]. It assumes abundance of information and digital networks, and views learning as the development and maintenance of networks of information, resources and contacts [14]. Primary activities in connectivist learning are [12]: i) aggregation, ii) remixing, iii) repurposing, and iv) forwarding of resources and knowledge.

Teaching in connectivist setting differs from common practices in distance and online education. In particular, teaching is focused on instructional design and learner facilitation, while the course content is created by course participants (i.e., learners and facilitators) [5, 6]. Kop et al. [15] therefore argue that the key to cMOOC success is a combination of teaching and social presence that enables an effective facilitation of learners' self-regulation of learning, which in turn leads learners to the accomplishment of worthwhile, personalized and authentic learning outcomes. Instead of being a distant "rock star" academic of xMOOCs [16] [p. 58], a teacher in cMOOC is expected to be a role model [14], and a discussion moderator rather than a tutor [12]. According to Kop et al. [15], instructors are "*aggregating, curating, amplifying, modeling, and persistently being present in coaching or mentoring. The facilitator also needs to be dynamic and change throughout the course*"[p. 89]. For this delegation of content creation from the instructor to the network, Yaeger et al. [9] emphasize the need for a strong core of active participants that would provide the critical mass of activity.

A typical design of a cMOOC assumes collaboration between course participants using various social media (e.g., blogs, Twitter, Facebook, Google+, RSS feeds and mailing lists) [17]. The use of particular tools and their affordances can directly influence and support the community formation [18], which is essential for learning within cMOOC environments. Twitter hashtags are probably the best example of technological affordances that can affect community formation [19]. However, the abundance and diversity of technology in cMOOCs is also a challenge [20]–[22], and a source of potential disconnect between the sub-communities in the course [14]. For example, a study by Mackness et al. [21] found that variations in the level of expertise and use of different platforms lead to the development of sub-communities which reduced possibilities for autonomy, openness and diversity. While cMOOC literature acknowledges the importance of technology for shaping learning experience, the effects of particular technologies are rarely discussed [3].

The cMOOC literature so far has mainly focused on descriptive methods for research and analysis of learning in a networked environment. Perhaps, the most comprehensive approach was applied in the study of Fournier et al. [23], who relied on counts of contributions/posts (e.g., Moodle discussion blogs, Twitter), survey, virtual ethnography, discourse analysis and educational data mining, in order to describe learning processes in the PLENK cMOOC. However, their discourse analysis relied on manual coding of messages, a highly time consuming process, while the quantitative methods applied (i.e., clustering and correlational analysis) did not provide a more detailed insight into the underlying learning processes. Although studies by Kop [9], and Yeager et al. [20] adopted social network analysis, the application was limited to the illustration of interactions within the course discussions. Finally, Wen et al.'s [24] study on discourse centric learning analyzed the association between learners' discourse and attrition in a MOOC, using the Latent Dirichlet allocation approach. However, they did not consider the principles of connectivism, nor did they consider different social media platforms.

2.2 Research questions

While the number of studies about MOOCs is growing [25], there have been very few studies that looked into the effects of particular choices of technology on shaping learning in cMOOCs. The exceptions are studies by Fini [17] and Mak et al. [26]. However, they primarily focused on quantitative analysis of interactions, media affordances and learning approaches, which did not provide insights into the content of learners' discussions. In our study, we wanted to examine learners' discourse in different social media that are typically used in cMOOCs – i.e., Facebook, Blogs and Twitter. The main objective was to obtain an insight into the topics that learners mentioned in their posts, and how these topics differ across different media. Accordingly we defined our first research question as follows:

RQ1: *Do topics discussed by learners differ across social media used in a cMOOC?*

In such a dynamic environment, where learners are encouraged to choose what they want to learn and make sense of the high volume of available information through sustained collaboration

[1] http://www.moocresearch.com

with other learners in a network, we were interested in examining the role of facilitators in shaping the discussions in the course. While the study by Skrypnyk et al. [27] identified the key role of a small number of active facilitators and technological affordances in shaping the information flow and formation of interest-based communities, it is still an open question how much these communities remain within the original course curriculum suggested by the instructors. Given that cMOOCs are typically organized as a series of online events led by respected facilitators in a particular domain [15], it seems reasonable to analyze how much influence those facilitators have on shaping the overall discussion between learners. This is likely related to the level of autonomy of learners, their self-regulation of learning, and their particular learning goals. Therefore, we defined our second research question:

RQ2: *To what extent do the readings suggested by the course facilitators shape the topics discussed by learners in social media in a cMOOC?*

We were also interested in examining whether the discussed topics stabilize over time or perhaps change in accordance with the changes in the course's weekly topics. This led us to our third research question:

RQ3: *How do topics discussed by learners change over time in a cMOOC across different social media?*

Finally, we aimed at providing a scalable approach for a comprehensive analysis of learners' discourse in cMOOCs. The study by Skrypnyk et al. [27] examined the use of particular Twitter hashtags over time and thus, to some extent examined the content of learner messages and their evolution over time. Still, our study provides a more comprehensive coverage of learners' generated discourse by investigating blog posts, Twitter messages and Facebook discussion messages.

3. METHODOLOGY

3.1 Study context

To get a better insight into the emerging topics in a cMOOC and answer our research questions (RQ1-3), we analyzed the content created and exchanged through social media in the scope of the 2011 installment of the Connectivism and Connective Knowledge (CCK11) cMOOC (http://cck11.mooc.ca/). The CCK11 course was facilitated through 12 weeks (January 17th – April 11th 2011), with the aim of exploring the ideas of connectivism and connective knowledge, and examining the applicability of connectivism in theories of teaching and learning. The topics covered throughout the course included: i) What is Connectivism?, ii) Patterns of Connectivity, iii) Connective Knowledge, iv) What Makes Connectivism Unique? v) Groups, Networks and Collectives, vi) Personal Learning Environments and Networks, vii) Complex Adaptive Systems, viii) Power and Authority, ix) Openness and Transparency, x) Net Pedagogy: The Role of the Educator, xi) Research and Analytics, and xii) Changing Views, Changing Systems. The course participants were provided with readings recommended by the course facilitators for each theme covered by the course (one theme per week). The facilitators encouraged learners to "remix" and share their new knowledge through various means including blogs, Twitter and Facebook[2]. The participants were also provided with daily newsletters that aggregated the content they created and exchanged through these blogs, tweets and Facebook posts. Content aggregation was done using gRSShoper. Finally, the course included weekly live sessions that were carried out using Elluminate.

3.2 Data Collection and Analysis

The overall process of data collection and analysis was done in several steps that are outlined below.

Collection of learners' posts and recommended readings. We relied on gRSShopper to automatically collect blog posts and tweets, while Facebook posts were obtained using the official Facebook API[3]. All posts were stored in a JSON format for further processing. Table 1 provides descriptive statistics of the posts collected. Besides posts, we also collected readings recommended by the course facilitators for each theme covered by the course. The recommended readings appeared in the course outline[4] for each week of the course.

Semantic annotation of learners' posts and recommended readings. Having collected learners' posts and recommended readings, the next step was to semantically annotate them, i.e., to associate their content with concepts that reflect the semantics of those posts and readings. To this end, we examined and tested several state-of-the-art semantic annotation tools, including TagMe[5], WikipediaMiner[6], Alchemy API[7], and TextRazor[8]. Based on the analysis of the annotations produced by the examined tools on a sample of the collected posts, and also based on the previous examinations of these tools reported in the literature (e.g., [28-30]), we made the following decision: short posts (tweets and Facebook messages) were annotated using TagMe, while Alchemy API was used for the annotation of longer posts (i.e., blog posts) and recommended readings. Both tools annotate content with Wikipedia concepts which made all the annotations consistent (i.e., based on the same concept scheme).

Since today's annotators mostly operate on English texts, we made use of a freely available language translation tool (Microsoft Translation API[9]) to translate non English posts (5% of our dataset) to English. Even though the resulting translations were not ideal, in most cases, we noticed that they preserved the gist of the original content.

Having inspected the annotations of posts and readings, we identified certain invalid concepts originating from the imperfection of today's semantic annotators. To reduce a potential negative impact on further analysis, we manually removed all concepts that were obviously erroneous (e.g., concept 'cable television' was identified as a disambiguation of the term 'networks', or 'environmentalism' was associated with '[learning]

Table 1. Descriptive statistics of the collected data: number of active learners, post counts (total, average, SD), and word count for each media analyzed

Media	Active participants	Post count	Average post count (SD)	Word count
Blog	193	1473	3.13 (4.80)	428626
Facebook	78	1755	5.03 (5.23)	67883
Twitter	835	2483	1.80 (3.85)	43180
Total	**997**	**5711**	**-**	**539689**

2 A complete list of the instructions provided to CCK11 participants is available at http://cck11.mooc.ca/how.htm

3 https://developers.facebook.com

4 http://cck11.mooc.ca/outline.htm

5 http://tagme.di.unipi.it/

6 http://wikipedia-miner.cms.waikato.ac.nz/

7 http://www.alchemyapi.com/products/alchemylanguage/concept-tagging/

8 http://www.textrazor. com/

9 http://msdn.microsoft.com/en-us/library/dd576287.aspx

environments'), as well as concepts that could not be considered valid in the context of our analysis (e.g., Lady Gaga's songs). Once we created a list of erroneous concepts, the removal was done automatically – before including a concept, we would ensure that the concept is not specified within the list.

Creation of concept co-occurrence graphs. The extracted concepts served as an input for the creation of undirected weighted graphs for each week of the course and each media analyzed (36 graphs in total). Aiming to identify the most important concepts and their connections, we created graphs based on the co-occurrence of concepts within a single post. For example, if concepts C1 and C2 appeared within the same post, the two concepts were included in a graph as nodes and the edge C1-C2 was created. Each edge was assigned a weight representing the frequency of co-occurrence of the two concepts.

Clustering of concepts into topics (concept clusters). To further analyze relationships between concepts in the constructed graphs, and extract clusters of concepts, we applied a modularity algorithm for community detection [31]. The initial analysis revealed a rather high number of clusters (over 50 on average, in case of Twitter graphs), with very few large groups and a significant number of small clusters (individual concepts or pairs of concepts). Therefore, we decided to extract the largest connected component in each graph, and use these components for cluster detection [36–38]. The size of the largest connected components used in the study varied from 88% to the size of the total graph in case of blogs, from 78% to 94% in case of Facebook, and from 52% to 86% of the total graph size in case of graphs extracted from Twitter.
In order to better understand emerging topics (i.e., clusters of concepts), we performed an in-depth qualitative analysis. We initially examined concepts within each cluster, aiming to reveal potential patterns that would provide description for the cluster analyzed. In cases where such a pattern could not be revealed, we focused on the content of the messages that these concepts were extracted from, to provide a better context for our interpretation.

Computation of graph metrics. The constructed graphs were analyzed using graph metrics that are commonly used for analysis of collocation networks [35]:

- *Graph density* – the ratio of existing edges to the total number of possible edges,
- *Weighted cluster density* – for each of the clusters we first calculated its graph density, and then calculated weighted average cluster density, where weights are cluster sizes. *Radius* – the minimum eccentricity among all nodes,
- *Diameter* – the maximum distance between two nodes,
- *Network centrality measures*, namely weighted degree (the count of edges a node has in a network, pondered by the weight of each edge) and betweenness centrality (the indicator of node's centrality in a graph).

The first three metrics were used to measure the level of coupling/spread of concepts (i.e., coherence) discussed in the analyzed posts, whereas the centrality measures served to measure the importance of individual concepts. Specifically, higher degree centrality should indicate concepts that are associated with many other concepts, while higher betweenness centrality could be seen as an indicator of concepts that could potentially "bridge" two or more topics [36]. Moreover, the selection of these metrics was motivated by the findings of contemporary research on automated assessment of learner generated content and information extraction. For example, Whitelock et al. [33] used keyword-based graphs for automated essay assessment and automated feedback provision. Their study showed that highly connected and dense graphs indicate better structured essays [37]. Building further on the research in computational linguistics, we expected that graphs with higher density would imply a more cohesive and coherent text [38]. Using the measure of degree, density, radius, and diameter, we aimed at examining whether and how the use of different media influences the "structure and cohesiveness" of the content being generated.

Computing similarity of posts as well as posts and recommended readings. To answer our research questions, we also needed to examine if there were topics of pertaining interest/relevance to learners, so that they kept discussing them even after the course progressed to other topics. To this end, for each social media analyzed, we computed the cosine similarity [39] between concepts discussed in each pair of consecutive weeks (i.e., concepts extracted from posts in the corresponding two weeks). In particular, we relied on a vector representation of the concepts discussed each week, and used the cosine similarity metric to compute similarity between concepts in two consecutive weeks. In a similar manner, we computed similarity between concepts discussed in posts and those discussed in recommended readings. In this case, the readings recommended for week *k*, k=1..11 were compared to posts in each succeeding week (k+1, k+2,…). The idea was to identify learners' interest in the course themes, based on the assumption that learners would discuss more topics that they find interesting/relevant.

4. RESULTS
In order to gain an initial insight into the topics discussed in each media channel, in Figure 1 we report the number of identified topics (i.e., concept clusters) identified and the most dominant topics for each media and each course week (Table 2, expressed as the percentage of the graph size, e.g., T1(45%)). We also examined the strength of relationships between concepts within the identified clusters (Figures 2 and 3); how concepts from different media relate to one another (Figure 4); the dynamics of concepts over the length of the course – whether and to what extent they changed from week to week (Figure 5 and Table 2), and how they relate to the recommended readings (Figure 6).

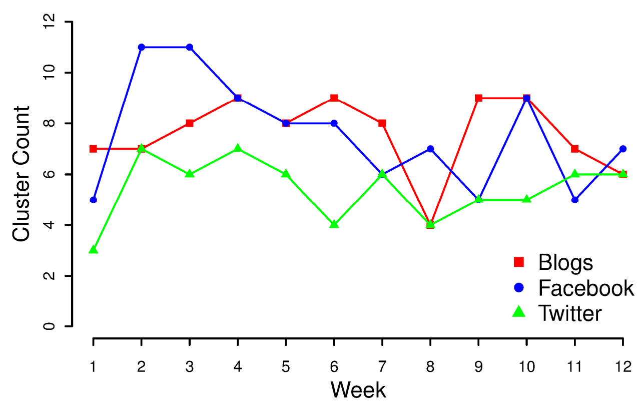

Figure 1. Topic (i.e., cluster of concepts) count per week per media

Figure 1 shows the number of detected topics (i.e., concept clusters) per week, for each media analyzed. Within the first half of the course, the highest number of topics was extracted from Facebook posts (except for week 1), while the messages exchanged on Twitter showed the lowest number of topics throughout the course.

Density of concept clusters for all analyzed social media follows quite a similar pattern throughout the course (Figure 2). Aiming to better understand the emerging concept clusters (i.e., topics), we calculated graph density for each individual concept cluster, per media and per week. It is interesting to note that the highest density among the media was observed in the first week of the course, for the concept clusters emerging from tweets. There are also two peeks where density increased notably; for blogs within the week 8, as well as by the end of the course in case of Facebook. These phenomena are analyzed in more details in the Discussion section.

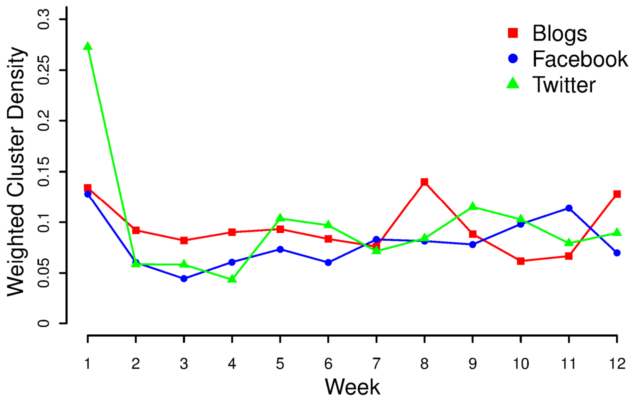

Figure 2. Average density of concept clusters per week and per media

Figure 3 further shows how concepts within topics (i.e., concept clusters) were coupled in terms of graph radius and diameter. The results show that concepts extracted from Facebook and blogs posts were more tightly coupled than those extracted from Twitter posts, which seems to indicate more homogeneous and related discussions overall on these two media. As the course progressed, concepts from tweets became more tightly coupled, while for Facebook and blog posts, the coupling of concepts remained approximately at the same level.

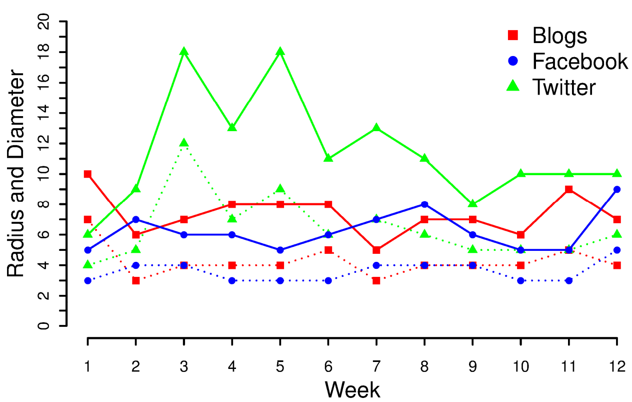

Figure 3 Radius (dotted lines) and diameter (solid line) of concept clusters measured per week and per media.

Figure 4 describes similarities between concepts discussed in each media. Comparison of concepts extracted from blogs and Facebook posts yielded the highest similarity over the 12 weeks of the course. On the other hand, concepts extracted from Twitter and blog posts showed the highest discrepancy throughout the course. It is also interesting to note the decline in similarity within the week 11, for each pair of media compared.

In order to further examine the dynamics of concepts being discussed, we calculated the similarity between concepts extracted from posts in each pair of consecutive weeks (e.g., for week 4, we calculated the semantic similarity of concepts from weeks 4 and 3). As a measure of semantic similarity, we calculated the cosine similarity between vectors of concepts for each pair of consecutive weeks. Figure 5 shows that in all media channels, the concepts discussed by learners remained rather similar from week to week. In case of Twitter posts, similarity between two consecutive weeks tends to increase over time (except for weeks 8 to 10), while in case of blogs and Facebook, we were able to observe a decrease over time.

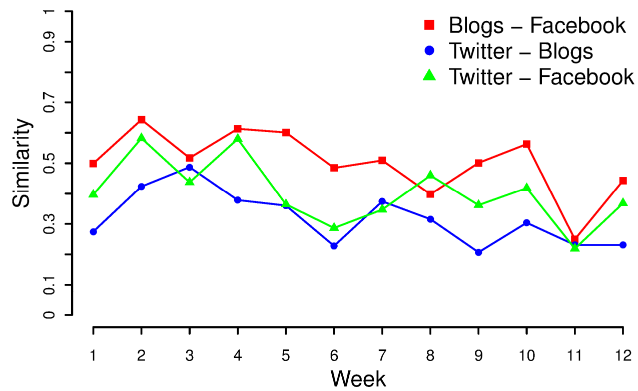

Figure 4. Similarity of concepts discussed in different media

We also analyzed semantic similarity between concepts extracted from posts exchanged on each media and recommended readings for i) the same week, and ii) all the previous weeks. For example, for week 7, we calculated similarity between concepts extracted from blogs, Facebook and Twitter in week 7, and concepts extracted from readings recommended in weeks 1 to 7. This analysis revealed a quite consistent pattern over the three media. Figure 6 shows that concepts extracted for each week, within all three media, were the most similar to the readings assigned for weeks 1-3, and 9. On the other hand, based on the extracted concepts, readings assigned for weeks 4 to 8 had the lowest similarity with posts from any of the course weeks. Moreover, among the three media analyzed, results show that Twitter posts (i.e., concepts extracted from Twitter posts) differed the most from the content presented in the readings for each week of the course, while blogs seemed to be the most similar to the readings.

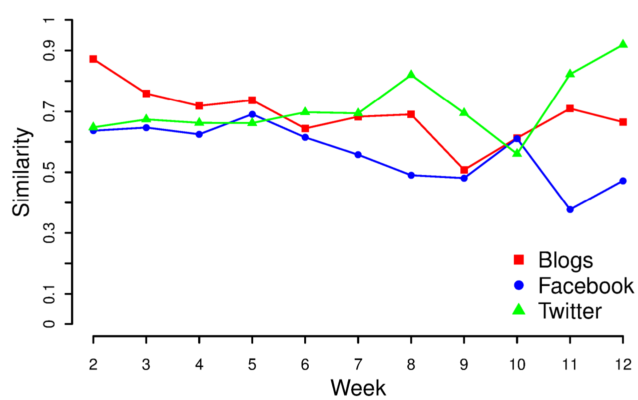

Figure 5. Similarity of concepts discussed in two consecutive weeks (per media)

Table 2 shows the top three topics (i.e., concept clusters) for each media and each week. Topics are ranked based on the number of concepts they consist of. For each topic, the table shows the top three concepts ranked based on their betweenness and degree centrality. Among those highly ranked concepts *connectivism, learning, e-learning, education, social media, and knowledge*, were most commonly represented within one of the three topics for most of the weeks, within each media analyzed.

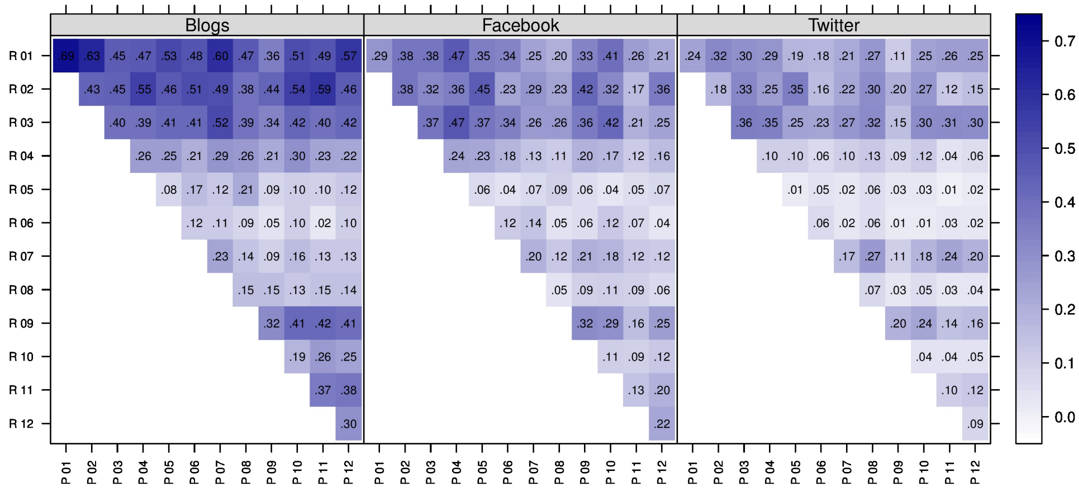

Figure 6. Similarity between weekly readings and posts from each week

An in-depth qualitative analysis of these results allowed us to provide a more detailed interpretation of the topics covered within each week, for each of the three media.

By analyzing topics identified in Twitter messages, we were able to identify the following five groups of topics:

- Within the first group of topics we recognized posts that are related to **sharing information** regarding the course, relevant publications, and other resources. These topics were indicative of weeks 1 to 3, as well as of weeks 7 and 11.
- The second group was based on topics related to connectivism **as a learning theory**. It is interesting to note that these topics were more frequent during the first four weeks of the course. Topics in this category included discussions on learning in networks (week 1); connectivism and its influence on instructional design (week 2); connectivism as one of the emerging learning theories (week 3); and unique characteristics of connectivism (week 4). Later in the course, topics such as connectivism as a learning pedagogy (week 8) received significant attention, as well as the potential influence of a connectivist approach to learning on changes in the role of instructional designers (week 9).
- The third group of topics was related to the application of **connectivism in practice.** The most notable points discussed included teaching foreign languages in connectivist settings and desirable competencies for teaching online (week 4); necessary skills for learning in networked learning environments (week 5); and the role of learners in connectivism and the importance of learning analytics (week 6). The topics belonging to this group received significant attention later in the course with the introduction of the concept "sharing for learning" in connectivism and available technologies for collaboration within a connectivist course (week 9). Finally, within the week 12 the role of connectivism in theory-informed research was also addressed.
- Within the fourth group of topics, **networked learning** and **establishing communities in networked learning environments** gained significant attention. Here, the course participants were interested in topics such as taking control of learning (weeks 2 and 3); networks and communities emerging from MOOCs (week 3); collaboration within networked learning environments (weeks 8 and 10); and design and delivery of social networked learning (week 12).

The final and the *largest set* of topics was primarily focused on **educational technology** and its application in various settings. The most indicative topics of this group are personal learning environments (weeks 5 and 6); social media in education (week 5); teaching with ICT and tools available (weeks 6 and 12); tools for learning and complex adaptive systems (week 7); integration of technological affordances into traditional classroom settings (week 8); challenges and best practices of educating teachers to use available technological affordances (week 9); and mobile (week 10) and blended learning (week 11).

Our analysis of topics detected in blog posts revealed topic groups similar to those observed in tweets, though with some observable differences:

- The first group of topics, similar to the one detected in Twitter messages, was about **sharing course resources**: information about the course and the readings (week 1), and the concept map of connectivism (week 11).
- The second group identified topics related to **MOOCs** in general: the concept of MOOC, previous MOOCs (e.g., PLENK, CCK08) (week 1), and how MOOCs affect learning in classroom settings (week 8). Although the topics from this group appeared throughout other weeks of the course, these topics were mostly discussed at the beginning of the course.
- The third group of topics received significant attention within the first five weeks of the course. This group was related to **connectivism as a learning theory**, and how connectivism relates to other learning theories. Course participants discussed the main characteristics of connectivism (weeks 1, 4, and 12) and relationships to other learning theories (week 5); validity of connectivism as a learning theory (week 2); teachers' role in connectivism (weeks 3 and 8); aspects of teaching English as a foreign language in connectivist settings (week 5); and about collective intelligence, constructivism, subjectivism and importance of interpretation (weeks 5 and 10).

Table 2. The number of exchanged posts and three most dominant topics (with the size as a percentage of all the clusters) for each week and each media; for each topic, the three most central concepts (sorted by betweenness and degree centrality) are given

	Twitter	Blogs	Facebook
Week 1	Total Topics: 3 Total Posts:30 T1 (45%): concept, substantial form, social T2 (27%): knowledge, open source, e-learning T3 (27%): connectivism, video, constructivism (learning theory)	Total Topics: 7 Total Posts:200 T1 (67%): learning, education, knowledge T2 (19%): twitter, concept, teacher T3 (6%): tag, critical thinking, website	Total Topics: 5 Total Posts:84 T1 (36%): connectivism, idea, learning T2 (25%): facebook, open source, uploading and downloading T3 (18%): information, paradigm, twitter
Week 2	Total Topics: 7 Total Posts:270 T1 (33%): connectivism, education, e-learning T2 (22%): employment, social network, thought T3 (22%): learning, concept map, instructional design	Total Topics: 7 Total Posts:159 T1 (35%): learning, knowledge, thought T2 (18%): argument, research, computer network T3 (18%): motivation, facebook, MOOC	Total Topics: 11 Total Posts:260 T1 (17%): twitter, facebook, quora T2 (17%): learning, tradition, employment T3 (15%): education, connectivism, knowledge
Week 3	Total Topics: 6 Total Posts:256 T1 (30%): connectivism, wikipedia, conversation T2 (26%): learning, knowledge, computer network T3 (15%): education, e-learning, stephen downes	Total Topics: 8 Total Posts:145 T1 (19%): thought, knowledge, social network T2 (17%): teacher, connectivism, information T3 (17%): mind, writing, metaphor	Total Topics: 11 Total Posts:189 T1 (21%): learning, thought, connectivism T2 (16%): linkedin, facebook, social network T3 (11%): knowledge, idea, object (philosophy)
Week 4	Total Topics: 7 Total Posts:236 T1 (23%): connectivism, education, constructivism (learning theory) T2 (20%): e-learning, social network, actor?network theory T3 (17%): learning, information age, theory	Total Topics: 9 Total Posts:160 T1 (25%): connectivism, knowledge, social network T2 (24%): theory, technology, time T3 (22%): thought, learning, education	Total Topics: 9 Total Posts:210 T1 (18%): knowledge, connectivism, social change T2 (18%): thought, e-learning, student T3 (16%): learning, education, skill
Week 5	Total Topics: 6 Total Posts:271 T1 (36%): e-learning, connectivism, bonk (video game series) T2 (24%): edtech, internet, english as a foreign or second language T3 (17%): education, educational entertainment, teacher	Total Topics: 8 Total Posts:182 T1 (27%): thought, theory, truth T2 (20%): sound, youtube, human T3 (18%): education, learning, connectivism	Total Topics: 8 Total Posts:269 T1 (24%): thought, knowledge, understanding T2 (23%): learning, education, student T3 (22%): connectivism, wiki, facebook
Week 6	Total Topics: 4 Total Posts:217 T1 (37%): connectivism, english as a foreign or second language, behaviorism T2 (32%): education, edtech, e-learning T3 (21%): collaboration, knowledge, thought	Total Topics: 9 Total Posts:109 T1 (18%): learning, education, psychology T2 (17%): feedback, connectivism, cognition T3 (15%): theory, book, internet	Total Topics: 8 Total Posts:144 T1 (20%): learning, thought, history of personal learning environments T2 (18%): knowledge, information, brain T3 (17%): diigo, blogger (service), tool
Week 7	Total Topics: 6 Total Posts:270 T1 (42%): connectivism, twitter, knowledge T2 (24%): edtech, e-learning, mind map T3 (14%): technology, complex adaptive system, department of education and communities	Total Topics: 8 Total Posts:122 T1 (22%): learning, education, knowledge T2 (17%): sense, idea, intention T3 (14%): complexity, understanding, human	Total Topics: 6 Total Posts:73 T1 (23%): education, knowledge, culture T2 (20%): twitter, united kingdom, facebook T3 (18%): information, employment, history of personal learning environments
Week 8	Total Topics: 4 Total Posts:207 T1 (37%): connectivism, writing, book T2 (30%): education, e-learning, edtech T3 (17%): social network, learning, power (philosophy)	Total Topics: 4 Total Posts:71 T1 (69%): learning, social network, psychology T2 (27%): research, neoplatonism, people T3 (3%): massive open online course, internet forum, beauty	Total Topics: 7 Total Posts:94 T1 (20%): knowledge, intelligence, information technology T2 (17%): education, rss, plug-in (computing) T3 (17%): research, social media, new media
Week 9	Total Topics: 5 Total Posts:156 T1 (42%): edtech, e-learning, web 2.0 T2 (33%): internet, connectivism, file sharing T3 (11%): learning, school, control theory	Total Topics: 9 Total Posts:87 T1 (26%): learning, education, hypothesis T2 (22%): thought, social group, happiness T3 (13%): skill, knowledge, literacy	Total Topics: 5 Total Posts:132 T1 (26%): education, student, technology T2 (22%): connectivism, knowledge, connectionism T3 (21%): learning, thought, object (philosophy)
Week 10	Total Topics: 5 Total Posts:160 T1 (38%): connectivism, computer network, pedagogy T2 (21%): e-learning, education, teacher T3 (19%): learning, MOOC, google apps	Total Topics: 9 Total Posts:111 T1 (27%): learning, education, educational psychology T2 (13%): facebook, google, twitter T3 (12%): truth, metaphor, behaviorism	Total Topics: 9 Total Posts:113 T1 (28%): learning, thought, connectivism T2 (22%): employment, student, collaboration T3 (19%): book, writing, child
Week 11	Total Topics: 6 Total Posts:228 T1 (36%): connectivism, social media, emergence T2 (25%): e-learning, edtech, education T3 (14%): learning, theory, information age	Total Topics: 7 Total Posts:76 T1 (22%): education, teacher, pedagogy T2 (21%): learning, psychology, science T3 (20%): thought, skill, concept map	Total Topics: 5 Total Posts:50 T1 (32%): knowledge, learning, quality (philosophy) T2 (21%): connectivism, thought, behaviorism T3 (18%): value (personal and cultural), wisdom, truth
Week 12	Total Topics: 6 Total Posts:182 T1 (31%): connectivism, web 2.0, networked learning T2 (28%): e-learning, education, edtech T3 (17%): learning, english as a foreign or second language, information age	Total Topics: 6 Total Posts:51 T1 (26%): thought, pedagogy, connectivism T2 (24%): learning, observation, education T3 (18%): writing, memory, attention	Total Topics: 7 Total Posts:137 T1 (22%): learning, research, connectivism T2 (20%): google, writing, English language T3 (18%): person, applied science, education

- **Networked learning** and learning in connectivist settings received the highest attention among the course participants who were using blogs as a communication medium. The main topics covered included complexity of learning in networks, professional learning and importance of motivation for learning in networked environments (weeks 2, 4, 7 and 12); tools for learning in networks and gathering information (week 2); groups versus networks in connectivist settings (week 3); importance of interactions, internal and external feedback for learning in networks (weeks 6, 7, and 10); the source of knowledge/intelligence in networks (week 8); the role of technology in mediating teachers' role in networked learning (week 11), and learning affordances in networked learning environments (week 9); and digital literacy (week 9) and conceptual models for learning in networks (week 12);
- Discussions about **online and distance education** represent the fifth group of topics. The most commonly discussed topics included e-learning in classroom settings (week 3); social media services and social media platforms in online and distance education (weeks 5, 7, 8, and 10); social networks, social groups, and emerging social communities in distance education (weeks 6 and 9); instructional design for alternative education (weeks 9, 10, and 12), and metrics for measuring learners' success in online and distance education (week 10).
- The final group of topics was concerned with **educational technology** and use of ICT in education. Virtual learning environments and their use in higher education (weeks 6 and 7), ICT for teaching foreign language (week 7), personal learning environments (week 8) and learning management systems in education (weeks 11 and 12), were most commonly discussed in blog posts.

According to our analysis, learners' messages exchanged on Facebook remained within similar general topics:

- **Available resources** and information about the **course content** were common topics within weeks 1, 2, and 12.
- Within the **connectivism as a learning theory** topic group, the course participants were discussing the idea of connectivism and its position in education (weeks 1 and 2); how connectivism was different from the paradigm "wisdom of crowds", collective and connective wisdom (weeks 3 and 11); the main challenges of new learning theories (week 7); origins of connectivism (e.g., connectivism as a connectionist approach to learning) (week 9), and how connectivism empowers learners to take responsibility for their learning (week 11).
- Similar to blogs, **networked learning** and **learning in connectivist settings** received the most significant attention. These topics were evenly distributed throughout the course, and included networked learning and affordances that foster learning and help development of digital literacies (weeks 1 and 2); nature of teaching and learning in connectivism (weeks 4 and 8); social networking groups and sharing information within networks (weeks 3, 5, and 10); assessment in the connectivist framework (weeks 10 and 11); and collaboration and cooperation in networks (week 11).
- As with other media analyzed, **educational technology** was quite significant topic starting from the week four of the course. Institutions of higher education and their view of the role of ICT in education (week 4); social media platforms and connectivism (week 5); personal learning environments and differences/similarities with learning management systems (weeks 6 and 7); tools for collecting, sharing and tagging resources (week 6); role of educational technology in teaching foreign languages (weeks 9 and 10); and ICT and intellectual ethics (week8), were the most prominent.
- Opposite to blogs where topics about online and distance education were quite prominent, within the Facebook communication channel, topics on **education** in general received more attention. Course participants were interested in advantages and disadvantages of formal and institutional learning (weeks 4 and 7); the role of scholars in digital environments (week 2); how we learn and where we are learning from (week 3); important characteristics and skills of learners that drive learning in general, and in connectivist settings (week 5), how to create knowledge from information (week 6).

5. DISCUSSION

5.1 Interpretation of results with respect to the research questions

Considering the subject of the course, it is not surprising that the most common topics covered within each media are related to *connectivism as a learning theory*, *networked learning*, *education* (in general, and online and distance education in particular), *skills for teaching/learning in networks*, and *educational technology*. However, concepts discussed within each topic differ to a certain extent. For example, among topics related to educational technology that were discussed in blog and Facebook posts, there was a topic covering the issues of teaching and learning with ICT. While the course participants, who discussed this topic through blog posts, were mostly focused on technological affordances in teaching foreign language, posts exchanged on Facebook discussed the same topic from the learners' perspective.

Regarding our first research question (RQ1), we found that except for the first week of the course and concepts extracted from Twitter, the topics learners discussed in their posts in all three media analyzed tended to follow a similar pattern. In particular, posts tended to cover a wide set of concepts that quite differed from one post to another (Figure 2). However, our findings also indicate that concepts extracted from Twitter posts less frequently co-occurred and were less tightly coupled within a topic than in case of blog and Facebook posts (Figure 2 and 3). It could be deduced that blog and Facebook allowed for writing more coherent posts. This confirms previous findings that social media vary in their affordances [40], in terms that certain social platforms allow for more elaborate writing on topics of interest. On the other hand, less coherent discourse might be an indicator of difficulties to form a learning community. Without a clear set of shared interest, it is unlikely that a community would emerge. Observing though the perspective of the three media analyzed, it seems that blogs and Facebook offer better opportunities for the community development.

As for our second research question (RQ2), we found that posts throughout the 12 weeks of the course mostly covered topics from recommended readings for the first three weeks. Within those three weeks of the course, readings included topics such as *connectivism as a learning theory*, *learning in networks*, as well as *learning in networks and connective knowledge*, which we identified as the most common topics in the analyzed posts. Moreover, Figure 5 shows that topics discussed within two consecutive weeks did not differ significantly, indicating that course participants tended to continue conversation on the topic of interest, rather than follow new themes introduced within the course. This suggests that those dominant themes are determined by groups of learners who engage collaboratively, rather than by the instructor. Therefore, we might conclude that our results support the main theoretical assumptions of connectivism [1] and are in line with the previous studies [8, 27]. More precisely, the learning process is not focused on transferring knowledge from the instructor to course participants, but rather on the connections and collaboration between learners [6], while learners also participate in content creation. Moreover Kop, et al. [15] and Skrypnyk et al. [27] confirmed that the information flow and knowledge building process also depend on those network-directed learners who are willing to engage into interaction with their peers and share knowledge among the network of learners. Therefore, it seems reasonable to conclude that learners engage into discussions with peers who share similar interests, thus framing the topics discussed within each media.

Finally, regarding our third research question (RQ3), our findings show that even though the count of topics identified within each week changed over time and differed among the media analyzed (Figure 1), the most dominant and high-level groups of topics (e.g., educational technology, networked learning) quickly emerged, and sustained throughout the course. More specialized concepts did change in each group of topic, since learners showed interests in various aspects of those topics (e.g., social network analysis, personal learning environments). However, overall they remained focused on the general groups of topics.

5.2 Limitations of this study

In order to address issues of internal and external validity of our findings, certain limitations need to be acknowledged. The main issues regarding internal validity originate in the process of data collection and concept extraction. In our study, we relied on gRSShopper for the automated collection of learners' blog posts, and copies of tweets. This source was used as by the time we collected data for the study (April-August 2014), several blogs were not available any longer. Likewise, due to the limitations introduced by the Twitter API, we were not able to obtain original tweets. Therefore, we turned to the posts available within the CCK11 newsletter. Second, we relied on Alchemy API and TagMe for the extraction of concepts from learners' posts and recommended readings. However, as stated in the Methodology section, these tools produced some erroneous concepts that we manually removed. This suggests that the extracted concepts might not fully and correctly represent the themes discussed in posts and readings. Finally, we relied on Microsoft Translate API in order to translate non-English posts (5% of all the collected posts), therefore the resulting translations depend on the quality of the API used.

Addressing issues of external validity is important from the perspective of generalizing our findings. Therefore, it is important to conduct a similar analysis within a different educational domain or course.

6. CONCLUSIONS

The reported study proposed a novel analytic approach that integrates tools and techniques for automated content analysis and SNA with qualitative content analysis. This approach was used for the exploration of topics emerging from the learners' discourse in cMOOCs, and offered an in-depth insight into the topics being discussed among course participants. Moreover, the proposed analytic method also allowed for validation of certain ideas of connectivism – e.g., learners were primarily focused on the course topics they were interested in, regardless of the topics suggested by the course facilitators, while the technology had a significant impact on how learners discussed certain topics [6]. Further, our approach might be suitable for analysis of different media used in cMOOCs, as one of the critical features. For such multi-media studies, it is essential to proceed to the analysis of actual content and discourse rather than just counts of the use (e.g., page hits) [41, 42]. This is necessary as different media have different affordances that can affect how processes of knowledge creation unfold in cMOOCs [18, 26].

Building a trustworthy community in diverse and large networks, as those emerging from cMOOCs, is recognized as one of the important challenges [26]. Being able to reveal topics discussed in different media and among emerging social groups might help learners to "bridge the social gap" and more easily reach groups with similar interests. On the other hand, our study also shows an overall low density of the analyzed concept graphs. This might be an indicator of low cohesion among the concepts used by learners [38], and low-to-moderate mutual understanding and consensus built within the entire network [37]. It seems that, at the network level, course participants could not find shared concepts of interests within those broader topics being discussed. In addition, our findings might indicate a lack of shared vocabulary or conceptual models, considering that people originated from different backgrounds and different cultures. However, a broad consensus of the entire network – per medium – might not be possible given the size and diversity in interests, background, and goals of the course participants. Perhaps, a better unit of analysis could be communities. For example, further research should create similar graphs for specific communities – e.g., such as those that emerged in the study reported in [27] – and analyze their cohesion, rather than the cohesion of the entire network. We would expect to reveal higher graph density, and more connected graphs, as indicators of higher level of shared understanding.

Our findings also indicate that several topics gained significant attention, while other course topics were not commonly discussed among learners. Therefore, the question is how facilitators and/or learners should proceed with regard to those less "interesting" topics? Given that learners choose what to learn in cMOOCs, should facilitators provide a better connection with those topics that were "more popular", or introduce "less popular" topics in different ways, or perhaps such findings could inform the course design, pointing out to the most important topics for the course participants?

Further research is also needed to examine how different social groups shape discussions and whether we can identify certain patterns in learners' approaches to course-related discussions, over various social media. For example, it would be interesting to analyze how social groups formed around certain topics evolve over time; are there groups that use various media to collaborate with their peers on a certain topic; and how much attention receive topics initiated by course facilitators, compared to topics proposed by learners.

7. REFERENCES

[1] G. Siemens, "Connectivism: A learning theory for the digital age," *Int. J. Instr. Technol. Distance Learn.*, vol. 2, no. 1, pp. 3–10, 2005.

[2] L. Pappano, "The Year of the MOOC," *The New York Times*, 02-Nov-2012.

[3] J. Daniel, "Foreword to the Special Section on Massive Open Online Courses MOOCs–Evolution or Revolution?," *MERLOT J. Online Learn. Teach.*, vol. 10, no. 1, p. 4, Mar. 2014.

[4] V. Kovanović, S. Joksimović, D. Gašević, G. Siemens, and M. Hatala, "What public media reveals about MOOCs?," *Br. J. Educ. Technol.*, vol. submitted, 2014.

[5] S. Bayne and J. Ross, "The pedagogy of the Massive Open Online Course (MOOC): the UK view," Curriculum (including learning, teaching and assessment) > Pedagogic approaches HEER000427, Mar. 2014.

[6] G. Siemens, "Teaching in Social and Technological Networks," *Connectivism*, 16-Feb-2010. .

[7] D. Gašević, V. Kovanović, S. Joksimović, and G.

Siemens, “Where is Research on Massive Open Online Courses Headed? A data analysis of the MOOC Research Initiative,” *Int. Rev. Res. Open Distance Learn. IRRODL*, vol. 15, no. 5, p. accepted for publication, 2014.
[8] R. Kop, “The challenges to connectivist learning on open online networks: Learning experiences during a massive open online course,” *Int. Rev. Res. Open Distance Learn.*, vol. 12, no. 3, pp. 19–38, 2011.
[9] C. Yeager, B. Hurley-Dasgupta, and C. A. Bliss, “cMOOCs and global learning: An authentic alternative,” *J. Asynchronous Learn. Netw.*, vol. 17, no. 2, pp. 133–147, 2013.
[10] M. De Laat, “Networked learning,” *Police Acad. Neth. Apeldoorn*, 2006.
[11] G. Siemens, *Connectivism: Learning as Network-Creation*. 2005.
[12] C. O. Rodriguez, “MOOCs and the AI-Stanford like Courses: Two Successful and Distinct Course Formats for Massive Open Online Courses,” *Eur. J. Open Distance E-Learn.*, vol. 2012, no. II, pp. 1–13, 2012.
[13] G. Siemens, “Connectivism: A learning theory for the digital age,” in *International Journal of Instructional Technology and Distance Learning*, 2005.
[14] T. Anderson and J. Dron, “Three generations of distance education pedagogy,” *Int. Rev. Res. Open Distance Learn.*, vol. 12, no. 3, pp. 80–97, 2010.
[15] R. Kop, H. Fournier, and J. S. F. Mak, “A pedagogy of abundance or a pedagogy to support human beings? Participant support on massive open online courses,” *Int. Rev. Res. Open Distance Learn.*, vol. 12, no. 7, pp. 74–93, 2011.
[16] J. Ross, C. Sinclair, J. Knox, S. Bayne, and H. Macleod, “Teacher Experiences and Academic Identity: The Missing Components of MOOC Pedagogy,” *J. Online Learn. Teach.*, vol. 10, no. 1, pp. 57–69, 2014.
[17] A. Fini, “The Technological Dimension of a Massive Open Online Course: The Case of the CCK08 Course Tools,” *Int. Rev. Res. Open Distance Learn.*, vol. 10, no. 5, 2009.
[18] A. Gruzd, B. Wellman, and Y. Takhteyev, “Imagining Twitter as an Imagined Community,” *Am. Behav. Sci.*, vol. 55, no. 10, pp. 1294–1318, 2011.
[19] L. Yang, T. Sun, M. Zhang, and Q. Mei, “We know what @you #tag: does the dual role affect hashtag adoption?,” in *Proceedings of the 21st international conference on World Wide Web*, New York, NY, USA, 2012, pp. 261–270.
[20] R. Kop, “The challenges to connectivist learning on open online networks: Learning experiences during a massive open online course,” *Int. Rev. Res. Open Distance Learn.*, vol. 12, no. 3, pp. 19–38, 2011.
[21] J. Mackness, S. Mak, and R. Williams, “The ideals and reality of participating in a MOOC,” in *Proceedings of the 7th International Conference on Networked Learning 2010*, 2010, pp. 266–275.
[22] M. Saadatmand and K. Kumpulainen, “Participants’ Perceptions of Learning and Networking in Connectivist MOOCs,” *J. Online Learn. Teach.*, vol. 10, no. 1, pp. 16–30, 2014.
[23] H. Fournier, R. Kop, and G. Durand, “Challenges to Research in MOOCs,” *J. Online Learn. Teach.*, vol. 10, no. 1, pp. 1–15, 2014.
[24] M. Wen, D. Yang, and C. P. Rose, “Linguistic Reflections of Student Engagement in Massive Open Online Courses,” in *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.
[25] T. R. Liyanagunawardena, A. A. Adams, and S. A. Williams, “MOOCs: A systematic study of the published literature 2008-2012,” *Int. Rev. Res. Open Distance Learn.*, vol. 14, no. 3, pp. 202–227, 2013.
[26] S. F. J. Mak, R. Williams, and J. Mackness, “Blogs and Forums as Communication and Learning Tools in a MOOC,” 2010, pp. 275–284.
[27] O. Skrypnyk, S. Joksimović, V. Kovanović, D. Gašević, and S. Dawson, “Roles of course facilitators, learners, and technology in the flow of information of a cMOOC,” *Br. J. Educ. Technol. Spec. Issue MOOCs*, p. submmited for review, 2014.
[28] M. Cornolti, P. Ferragina, and M. Ciaramita, “A Framework for Benchmarking Entity-annotation Systems,” in *Proceedings of the 22Nd International Conference on World Wide Web*, Republic and Canton of Geneva, Switzerland, 2013, pp. 249–260.
[29] L. Jean-Louis, A. Zouaq, M. Gagnon, and F. Ensan, “An Assessment of Online Semantic Annotators for the Keyword Extraction Task,” in *Proceedings of the 13th Pacific Rim International Conference on Artificial Intelligence*, Gold Coast, Australia, 2014.
[30] J. Jovanovic, E. Bagheri, J. Cuzzola, D. Gasevic, Z. Jeremic, and R. Bashash, “Automated Semantic Annotation of Textual Content,” *IEEE IT Prof.*, vol. (to appear), 2014.
[31] M. E. Newman, “Modularity and community structure in networks,” *Proc. Natl. Acad. Sci.*, vol. 103, no. 23, pp. 8577–8582, 2006.
[32] V. Batagelj and A. Mrvar, *Pajek—analysis and visualization of large networks*. Springer, 2004.
[33] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, “Measurement and Analysis of Online Social Networks,” in *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, New York, NY, USA, 2007, pp. 29–42.
[34] W. Lee, J. J. Lee, and J. Kim, “Social Network Community Detection Using Strongly Connected Components,” in *Trends and Applications in Knowledge Discovery and Data Mining*, Springer, 2014, pp. 596–604.
[35] F. Boudin, “A comparison of centrality measures for graph-based keyphrase extraction,” in *International Joint Conference on Natural Language Processing (IJCNLP)*, 2013, pp. 834–838.
[36] N. Schluter, “Centrality Measures for Non-Contextual Graph-Based Unsupervised Single Document Keyword Extraction,” in *Proceedings of TALN 2014 (Volume 2: Short Papers)*, Marseille, France, 2014, pp. 455–460.
[37] D. Whitelock, D. Field, S. Pulman, J. Richardson, and N. van Labeke, “Designing and testing visual representations of draft essays for higher education students,” in *Discourse-Centric Learning Analytics (DCLA) Workshop, LAK*, 2014, vol. 14.
[38] A. C. Graesser, D. S. McNamara, and J. M. Kulikowich, “Coh-Metrix providing multilevel analyses of text characteristics,” *Educ. Res.*, vol. 40, no. 5, pp. 223–234, 2011.
[39] A. Huang, “Similarity measures for text document clustering,” in *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand*, 2008, pp. 49–56.
[40] R. F. Schmid, R. M. Bernard, E. Borokhovski, R. M. Tamim, P. C. Abrami, M. A. Surkes, C. A. Wade, and J. Woods, “The effects of technology use in postsecondary education: A meta-analysis of classroom applications,” *Comput. Educ.*, vol. 72, pp. 271–291, 2014.
[41] H. Fournier, R. Kop, and H. Sitlia, “The Value of Learning Analytics to Networked Learning on a Personal Learning Environment,” in *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, New York, NY, USA, 2011, pp. 104–109.
[42] P. Goodyear, *Advances in Research on Networked Learning (Computer-Supported Collaborative Learning, V. 4)*. Norwell, MA, USA: Kluwer Academic Publishers, 2004.