

Linking Online Identities and Content in Connectivist MOOCs across Multiple Social Media Platforms

Rafa Absar

The iSchool at The University of British Columbia
Suite 470-1961 East Mall, Vancouver, BC, Canada, V6T 1Z1
rafa.absar@mail.mcgill.ca

Anatoliy Gruz

Ted Rogers School of Management, Ryerson University
350 Victoria Street, Toronto, ON, Canada, M5B 2K3
gruzd@ryerson.ca

Caroline Haythornthwaite

The iSchool at The University of British Columbia
Suite 470-1961 East Mall, Vancouver, BC, Canada, V6T 1Z1
c.haythorn@ubc.ca

Drew Paulin

School of Information, University of California, Berkeley
102 S Hall Rd #4600, Berkeley, CA 94720, United States
drew.paulin@ischool.berkeley.edu

ABSTRACT

In this paper, we examine how multiple social media platforms are being used for formal and informal learning by examining data from two connectivist MOOCs (or cMOOCs). Our overarching goal is to develop and evaluate methods for learning analytics to detect and study collaborative learning processes. For this paper, we focus on how to link multiple online identities of learners and their contributions across several social media platforms in order to study their learning behaviours in open online environments. Many challenges were found in collection, processing, and analyzing the data; results are presented here to provide others with insight into such issues for examining data across multiple open media platforms.

Categories and Subject Descriptors

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

General Terms

Measurement, Human Factors.

Keywords

Social Media, Learning, Connectivism, Social Networks, MOOCs.

1. INTRODUCTION

Emergent internet technologies have greatly influenced learning environments, and the roles and behaviours that both learners and educators enact in pursuit of creating and sharing learning experiences. Social media are at the forefront of this transformative shift, bridging the social relationships and communities in which learners participate, and the discovery, sharing, filtering, and often co-construction of knowledge and

information that is a principal aspect of the online world. Whether learners are in a formal learning context or are informally seeking out new learning experiences under their own direction, they are starting to turn to various social media platforms to seek out information through individuals and communities with a shared learning interest. However, with the turn to social media, both instructors and learners are yet again challenged to develop learning new practices for constructing these collective learning spaces. Learners are often leveraging a number of different social media platforms, communities, and networks towards the achievement of their personal learning objectives and goals. The collaborative, social nature of learning that takes place engenders a pedagogical approach that is based on forming connections, the exchange of resources, and harnessing information flow across networks and social media platforms that allows for emergent, self-directed learning to flourish [15].

A question that arises is what work do these media accomplish in support of learning? But, to pursue this question, it is necessary to understand the way media are being brought into learning, and to determine means of collecting and managing that data. Connectivist Massive Open Online Classes – cMOOCs – combine both the pedagogical intent of use of multiple media, and the user-directed selection of media. Thus, examining media use in conjunction with these MOOCs has outcomes that can be beneficial for learning outcomes in such MOOCs, and as feedback to instructors on cMOOC activity. Yet, as found in our work, the potential media multiplexity of such MOOCs presents a number of challenges with respect to data collection, identity resolution and analysis.

The overarching goal of our research as a whole is to develop and evaluate methods for learning analytics to detect and study learning processes that might be happening across multiple social media platforms in both formal and information settings. This paper describes an early step in this research, focusing on identifying challenges and solutions associated with the collection and processing of social media data from multiple platforms. In our work, we found a major task – and thus also a challenge for analysis of multiple media use – to be assembling the whole of a learner profile, i.e., identifying individuals across the names and identifiers used on different platforms, and attributing to them the contributions made on these multiple platforms.

The paper starts with a review of relevant literature about the multi-platform landscape of social media and learning, and discusses areas of interest that arise when people rely on multiple social media platforms to support learning and teaching (or what we call “social media multiplexity”). We also include some recent data obtained from a questionnaire about social media use by instructors in higher education [28]. Then we turn to the MOOC data issues, describing first pedagogical structures for multi-media use outlined for the two connectivist MOOCs examined, and then the array of data and identity resolution issues we faced when processing of data from these two MOOCs. The paper concludes with implications for future work in this area.

2. SOCIAL MEDIA AND LEARNING

Learning in new networked, mediated spaces is socially embedded. It is tied to the interests of the learner, the multiple, often overlapping, social spheres that the learner interacts with through various social media and communities, and the relationships and social contexts that are formed through such interactions. Learning through participation and social construction of understanding and meaning, along with the development of understanding how to be a member of a knowledge community, are core aspects of what has been termed Learning 2.0 [3]. Stemming from Learning 2.0 practices is a culture of freely creating and sharing content, along with opportunities for groups and crowds to share and participate in social learning activities, that is driving a change in who learns what, from whom, and via what means [19]. The roles that social media play in learning experiences are key to this development.

2.1 Formal and informal learning contexts

Higher education faculty are recognizing the value that social media can leverage in their curriculum, with over one third of teaching faculty in the US using some form of social media in their courses, and adoption rates of social media as high as 80% among university classrooms in the US [18]. A perusal of recent EDUCAUSE Studies of Undergraduate Students and Information Technology [6, 22] indicates that social media are both being formally integrated into institutional academic learning experiences, and are also being used informally by students to supplement their learning experiences. This allows students to reach wider social networks via social media while simultaneously “meeting the student population where it lives: i.e., online, in social networking sites and in the microforms of communication adopted in Twitter” and other popular online platforms [10].

Results from a recent survey by our research group provides further information on social media use by instructors [28]. In response to a questionnaire, the 333 respondents, showed a wide use of social media as presented in Figure 1. While ‘consumption’ of media is more prevalent than ‘contribution’, these early social media adopters are still active contributors. Moreover, they overwhelmingly find and use media from *outside* their institutional learning management system (Figure 2). Thus, even outside the MOOC environments, the greater use on non-institutional platforms increases the difficulty of determining the full extent of learners’ media-based contributions.

Whether instructor-led or learner initiated, learners are using various forms of social media to bridge the gap between in-school and out-of-school learning. Use enables discovery of connections between traditional curricula, personal interests, and online communities [14], and connections to peers, communities, and resources across time and space [5]. Using resources in a self-directed manner enables the personalization of learning

experiences, and the construction of personal learning environments [7, 17, 21].

Figure 1: Social media use by instructors (ordered by frequency of ‘Contribute’ response)

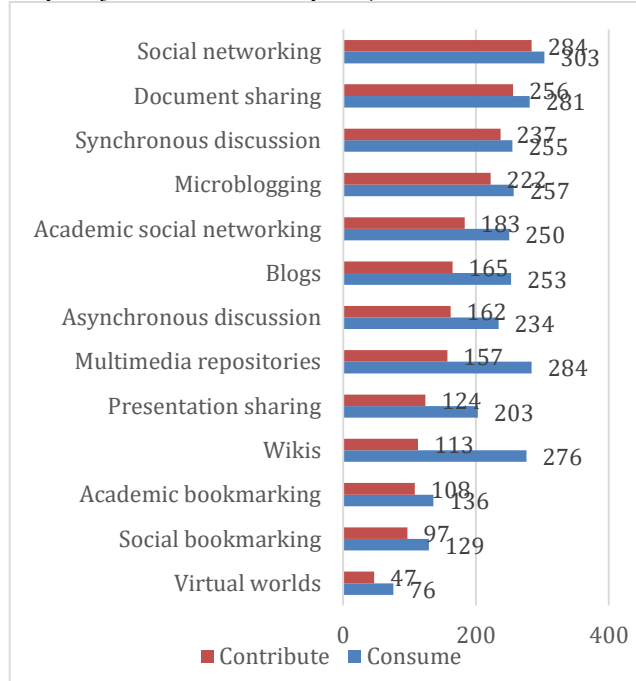
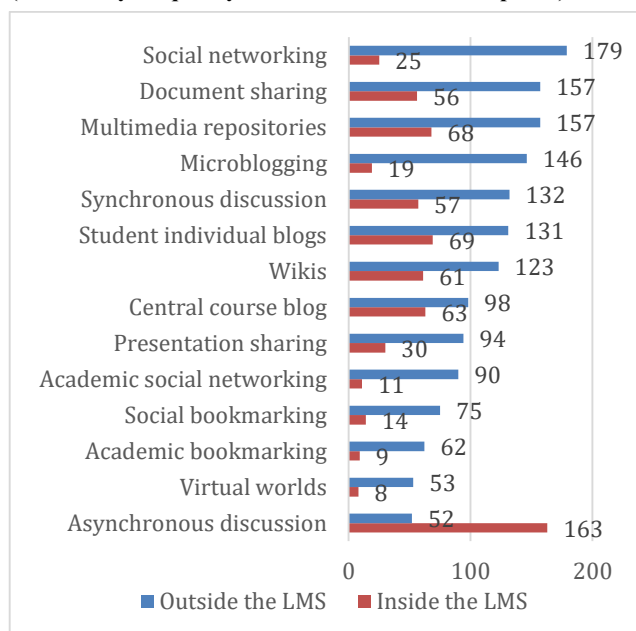


Figure 2: Use of social media inside and outside the LMS (ordered by frequency of ‘Outside the LMS’ response)



2.2 Social media multiplexity

Collaboration within one or more communities across a network supports the “social fabric” of learning [24]. Engagement in online communities benefits one’s learning through the establishment of trust and confidence in a learning community, along with one’s ability to create, re-use, and reimagine discourse and artifacts across communities in one’s learning network [15].

For these reasons, participation in a number of platforms and communities allows learners to approach concepts and ideas from various perspectives and purposes, and exposes them to a variety of viewpoints to consider across a number of communities of learning. This connectivist approach to learning – where learners negotiate and construct meaning and knowledge across a network of learners, platforms, and information sources – emphasizes the interconnected nature of learning. Learners develop their sensemaking abilities by relating knowledge fragments across a variety of environments, within a large pool of collective knowledge [20].

Users associate different aspects of their lives, or even different facets of their identities, with particular social media platforms. For example, while LinkedIn and Facebook share many functional commonalities as social networking sites, users leverage them in very different ways and share very different types of information with their networks on either of these platforms. Xu et al. [25] conducted a large-scale user alignment study that mapped over 21 million users across 6 social media platforms (Twitter, Tumblr, Wordpress, Blogger, Instagram, and Facebook) in order to quantify the level of user engagement on different combinations of these platforms, and to identify the number of overlapping users. The results indicate that the majority of users use a combination of these platforms.

2.2.1 Challenges for users and researchers

From a user's perspective, managing information sharing across various platforms and personal preferences specific to each can be strenuous. Vu et al. [23] propose a system that aggregates the social data of an individual subscribed to multiple platforms, and identifies and filters relevant information to be shared on a particular platform with a particular subset or community of that users' platform-specific social network. For learners, particularly those engaged in connectivist MOOCs, the management of knowledge across several platforms and communities can be a difficult task with a great deal of cognitive overhead. To address this issue, knowledge management systems and web applications have been proposed and developed in order to aggregate information, resources, services and people across a number of platforms towards the facilitation of a personal learning environment [2, 9].

For those researching social media use, media multiplexity offers complex challenges to overcome. For example, connecting a user's identity across multiple social media is a difficult challenge for researchers to address, but one that must be attended in order to gain a deeper understanding of learning processes that are occurring within and surrounding social media use [10]. Attempts to address this challenge are currently underway. For example, Zafarani and Liu [26, 27] describe an approach that finds and maps a user's identity across social media platforms by focusing on unique behavioural patterns exhibited by a single user across platforms, informational redundancies offered by the user across platforms, and machine learning towards effective user identification.

2.2.2 Multiplexity: insights and research questions

Along with challenges, media multiplexity offers insight on social media use. Because of the combinatory use of platforms, data from one platform can provide insight on how another platform is used. The combination of multiple social media datasets allows insights that go beyond what could be obtained by looking at one individually. Abisheva et al. [1] combine user-centric data from Twitter with video-centric data from YouTube to construct a descriptive analysis of the demographic and behavioural features

of YouTube usage through video links shared in Twitter. This approach provides a deeper understanding of who watches and shares what on YouTube, and how quickly videos are shared with others in users' networks across both platforms.

Media multiplexity also provides a number of interesting questions for researchers to address. For example, how do the fluencies, literacies, and sociotechnical practices that learners master for each specific platform or community influence their interactions on other platforms? What patterns of learning processes and sequences can be mapped and attributed to which sets of social media platforms? Cross-media literacy and practices become important as information and communication flow across media and technology, and the boundaries on what form is appropriate, when and where, are highly porous and also unpredictable [12, 13]. This leads to further questions about the formation and maintenance of interpersonal relationships across media multiplexity, and multiplex constructs of collaboration and community. This new landscape of social media enacts multi-modal learning, requiring the development of skills and competencies in communication and conversation, information retrieval and contribution on and across multiple platforms, both singly and in combination. It also engages learners as members of crowds and communities with multiple stakeholders, from novice to expert, with and without defined roles, across networks constituted in formal and informal contexts focused towards work, socialization, play, and learning [13].

2.2.3 Personal learning environments and cMOOCs

As learners participate across content, contexts, and dialogues through various social media platforms, they construct Personalized Learning Environments (PLE) that allow them to create, organize, and share content with other learners in their network [5, 16]. PLEs are "tools, communities, and services that constitute the individual educational platforms that learners use to direct their own learning and pursue educational goals" [8], and may be interpreted as the representation of a learner's informal learning activity that takes place online, across a number of different platforms and spaces [5]. Connectivist MOOCs are learner-centered, offering participants self-directed social learning opportunities across a number of social media [7, 21]. In cMOOCs, we see the convergence, interconnection, and often coordination of PLEs of many learners and educators across a network of social media platforms and communities.

3. MULTI-PLATFORM cMOOCs

The two courses from which data was extracted and on which this paper reports on are the connectivist courses CCK11 (Connectivism and Connective Knowledge 2011) and Change11. CCK11 ran for 12 weeks (January to April 2011) while Change11 ran for a longer period of 35 weeks (September 2011 to May 2012). Both courses discussed the idea that knowledge is networked and learning is distributed. The intent of a cMOOC is to encourage connection between people, concepts, writings, resources, and move. Discussions and learning processes in these courses were supported through the following four tasks:

- 1) Aggregate: Participants were given access to a wide variety of things to read, watch or play with.
- 2) Remix: Participants were then encouraged to keep track of their in-class activities using blogs or other types of online posts.
- 3) Repurpose: Participants were asked to not just repeat what other people have said, but create their own content.

- 4) Feed Forward: The last step was to share their work with other people in the course or outside the course to spread the networked knowledge.

Course resources were provided using gRSShopper and online seminars delivered using Elluminate. The courses, however, were not restricted within a single platform or environment, and hence the content was distributed across the web. Participants were free to use a variety of technologies for sharing and participating in the course. To keep track of their learning and sharing content, participants were encouraged to create blogs using any blogging service, including blogger.com or wordpress.com, use del.icio.us, discuss on Google groups forums, tweet about items on Twitter, or use anything else such as Flickr, Second Life, Yahoo Groups, Facebook, or YouTube.

To be able to keep track of their content, participants were asked to use the #cck11 or #change11 tags in whatever content they created and shared. These tags were used to recognize content related to the courses using aggregators. The aggregated content was then displayed in an online “newsletter”, which was created everyday to highlight some of the new content posted by learners.

To collect data, we scraped the archives of the daily newsletters for each course, and automatically extracted information on four types of data – Twitter messages, discussion threads, blog posts, and comments on blogs. The most popular platform (in both CCK11 and Change11) that generated the most number of posts was Twitter, followed by blogs (see Table 1).

Table 1: Number of posts on each platform

Platform	CCK11	Change11
Twitter posts (tweets)	1722	5665
Blog posts	812	2486
Blog comments	306	134
Discussion thread posts	68	87

In our research, we are interested in examining why and how online learners choose one platform over another or whether they use multiple platforms at the same time during the course of their class participation. We would also like to know whether and how each of the available platforms contributes to one’s individual and collaborative learning. However, before we could answer these fundamental questions, first we need to retrieve and prepare our data to ensure their consistency and quality, as discussed below.

4. DATA CHALLENGES

Several research and technical challenges come up when collecting and processing data from cMOOCs, primarily because they do not use a single, centralized platform to support class interactions. This section discusses these challenges and offers some solutions to be used in our future work.

4.1 Data collection issues

We developed a custom script to collect (scrape) data from a publicly available archive of daily newsletters that included snippets of participants’ blog posts (with the link to the original posts), comments left on each blog post, threaded discussions and Twitter messages. For these particular classes, class interaction data were archived and remained web accessible even after the end of the classes. However, for other classes where messages are not archived, the use of a real-time data collection tool might be required. Furthermore, since only Twitter messages that included class-specific hashtags (#CCK11 or #Change11) were recorded in

the daily newsletters, we likely missed any direct replies among class participants that did not use these hashtags.

Since only snippets of blog posts were available in the daily newsletters, we attempted to “follow” links to the original blog posts to retrieve them as well. During this process, we encountered a number of technical issues. First, since blogs were hosted on different servers (independent from the class website), some URLs to blogs maintained by students have expired, leading to timeouts or “server not found” messages. Some domains have disappeared completely by the time of our data collection, while others are now password-protected. Some URLs launch modal dialogs (pop-ups) that need to be processed by the scraping tool accordingly to prevent the collection process from crashing. Some web pages launch scripts that may also crash the collection tool.

As part of this process, we also identified some issues related to the encoding of several webpages. The target HTML tags used for scraping the webpages need to be set for every host domain. This is less of a problem if a lot of the participants use the same domain (e.g. Wordpress or Blogspot), but more of a problem if they use self-hosted blogs on unique domain names. Some pages (particularly the instructor’s pages) did not contain tags with class or ID attributes that make tag targeting easier. Finally, a few pages used non-standard tag formats, which required either filtering or using different parsing configurations that slowed down the data collection.

Other issues arose from the course material itself, since this can be presented and stored in a variety of formats. For example, some posts and webpages contained mostly images or video, with little or no explanatory text to scrape. Some course events were live seminars (presented on Elluminate) and did not include transcripts that can be scraped. This suggests that in order to fully capture all aspects of the class, one might want to consider using additional tools designed to retrieve and analyze multimedia content such as images and videos. In our case, since the majority of in-class interactions were text-based and due to the lack of effective tools to handle multimedia content automatically, we decided to focus only on the retrieval and analysis of text-based messages and exclude multimedia content.

The fact that participants were located around the world added another challenge to the pre-processing stage. More specifically, some of their posts were not in English and hence they required special consideration prior to the analysis stage. For example, one option would be to translate any non-English text automatically. For the purposes of our exploratory study, we decided to exclude non-English posts since they represented only a small portion of all posts.

Data collection also requires a significant amount of cleaning after extraction was complete. Some newsletter elements were redundant (e.g. point to the same blog entry) and had to be detected and removed. Discussion threads as stored in the course archive page only included replies to the original post, but unfortunately did not include the original post itself. This is further complicated by the fact that although all blog postings were RSS’d to the course page (by request of the course instructors), they were not archived, so we have to try to scrape the original blog pages themselves (i.e., if they are still available).

Finally, we noticed that some newsletter pages were missing from the archive. These are just some of the most common technical challenges that we encountered and that would need to be addressed by any similar data collection protocol.

4.2 Identity resolution

To analyze cross-platform data holistically, we need to be able to combine the data across all platforms. For this, we need an effective way to distinguish between the same participants across different platforms. The challenge is that participants may use different usernames for each platform, e.g., one for Twitter, one for blogs, etc. This kind of identity resolution has often been addressed using computational linguistics and machine learning techniques [11] and can be separated into two primary tasks: *coreference resolution* (resolving single identities across multiple platforms) and *alias resolution* (identifying two or more people with the same name or alias across platforms).

The quality of identity resolution approaches can be improved by also matching any metadata available about the users. Researchers often rely on users' profile information to find same users across different platforms [4]. However, in our case, cMOOCs participants could (and some did) remain entirely anonymous when participating in the class. As a result, in many instances, we simply did not have enough information to match users across platforms automatically. Therefore, to achieve the highest possible level of data, we have taken a manual brute-force approach of matching usernames across platforms. This also helped us identify challenges if trying to automate this process in the future.

During this process, every username from each platform was matched and cross-referenced with usernames from other platforms and any matching ones were grouped together. After exact matches, partial matches were also looked at – for example, jdoe and johndoe. Optional fields of First and Last names helped to determine if a slightly different username across two platforms may be the same too. For e.g., jdoe and johndoe are likely to be the same person if no other Doe's with first name starting with J are present in any of the platforms. This method is repeated for each platform until completed.

Through this process, we identified the unique users (aliases) who posted in each platform (see Table 2). We were also able to identify the users that posted in more than one platform (see Table 3). Although the number of users who posted across three or more platforms is small, a reasonable number of participants posted in at least two platforms. It should be noted that in Change11, out of the 103 users who did post in two platforms, the preferred media for posting or sharing for 93 of them were blogs and Twitter.

Table 2: Number of unique users who posted on each platform

Platform	CCK11	Change11
Twitter	145	794
Blog	105	278
Blog comments	56	27
Discussion thread posts	18	17

Table 3: Number of users who posted in more than 1 platform

No. of users posting in...	CCK11	Change11
4 platforms	2	3
3 platforms	10	5
2 platforms	32	103

Although this is a promising approach, it is time-consuming and may miss identifying single identities or erroneously group two separate identities as one. A faster, more reliable and real-time identity resolution method is required.

5. CONCLUSION AND FUTURE DIRECTIONS

The paper reported on the first stage of a larger project that strives to develop learning analytics methods to support assessment of collaborative learning in social media; specifically, when studying social media-based learning environments that rely on multiple platforms ("social media multiplexity"). As the literature review section suggests, people are often engaged in learning processes (especially informal learning) on multiple social media platforms. In this context, our broad agenda is to determine whether and what types of learning occur on different social media platforms and how learners choose what platform to use and for what purposes. This current paper focused on the first stage of this research, which is to determine technical and logistical challenges (and possible solutions) associated with the collection of social media data and identity resolution across multiple platforms.

As a case study, we examined two cMOOC-type courses that relied on blogs, Twitter, and discussion forums to organize the class. We found that most of the technical challenges during the data collection process were access-related, and primarily due to the disappearance of old online resources and links. This was usually because we worked with so-called "legacy" data that remained publicly available once the classes were over. This suggests that one way to limit this type of issues is to implement live collection of class content and interactional data as it is being generated by online participants. If this is not feasible, we believe that our experience working with the archived social media data, as reported in this paper, will be a valuable starting point to those who are working on a similar project.

Since people may use different aliases on different social media platforms, another critical issue in this line of research is how to identify and "follow" the same individual across multiple platforms. Although the previous research has identified a number of automated approaches to address this issue, including techniques such as *coreference* and *alias resolution*, our manual analysis of users' accounts across multiple platforms for the two classes suggests that the class participants mostly relied on a single platform. Only a small percentage of users posted to multiple platforms during the class. Furthermore, since one of our primary goals is to assess collaborative learning, we are especially interested in analyzing interaction data among learners. But among the four different ways of interacting with others in the class (blog posts, blog comments, tweets and discussion threads), Twitter was the single, most popular platform for discussion. And even though blogs (specifically blog posts) were the second largest content generators after Twitter; our manual review of the blog posts revealed that they were primarily used to take notes and write reflection-type pieces, and they were not used to interact with one another. In sum, we started this project with the expectation that we would need to identify and resolve online identities across multiple platforms; but in reality, both classes primarily relied only on Twitter for user-to-user interaction. This is despite the fact that class participants were encouraged to use a wide range of social media platforms. Therefore, we conclude that for the two classes in question, the need for identity resolution is very low. We hypothesize that we might find a similar pattern in other cMOOC-type courses where only a few of the most active users rely on two or more social media platforms for class participation. Our future work will test this hypothesis on more datasets. If it holds, the main implication is that there might not be a need for resource-intensive identity resolution algorithms when studying social media multiplexity and that we might be able to

analyze each platform independently from other platforms, at least from the perspective of identifying overlapping communities of users across platforms.

The limitation of our data is that we cannot tell whether people are actually reading some of the content on platforms where they are not active. In our future research, we plan to conduct the content analysis of messages posted on different social media with the goal to identify the level of cross-referencing across platforms. For example, when a student promotes a blog post on Twitter or starts a discussion thread to discuss a resource mentioned by another class member in a blog post. Our next step is also to analyze emerging communication networks on each of the platforms to characterize and evaluate the types of connections and communities (or crowds) formed across participants and how they relate to the nature and type of the platform used. Finally, for few participants who did engage on multiple platforms, we would like to know whether their position in communication networks changes depending on the platform and why. Answers to these questions will help us determine the types of learning analytics techniques that might be useful in studying collaborative learning in such environments.

6. ACKNOWLEDGMENTS

This work was supported by the Social Sciences and Humanities Research Council of Canada (SSHRC) grant awarded to Dr. Anatoliy Gruzdt and Dr. Caroline Haythornthwaite. Initial results of this research appeared in the Proceeding of the 2015 International Conference on Learning Analytics and Knowledge.

7. REFERENCES

- [1] Abisheva, A., Garimella, V.R.K., Garcia, D., et al. 2014. Who watches (and shares) what on youtube? And when? Using Twitter to understand Youtube viewership. In *Proceedings of the 7th Int. Conference on Web search and data mining*. ACM, 593–602.
- [2] Bogdanov, E., Limpens, F., Li, N., et al. 2012. A social media platform in higher education. *Proceedings of the IEEE Global Engineering Education Conference*, 1–8.
- [3] Brown, J.S. and Adler, R.P. 2008. Minds on Fire: Open education, the Long Tail, and learning 2.0. *EDUCAUSE Review*. 43, 1, 16–32.
- [4] Carmagnola, F., Osborne, F., & Torre, I. 2010. User data distributed on the social web: how to identify users on different social systems and collecting data about them. *Proceedings of the Int. Workshop on Information Heterogeneity and Fusion in Recommender Systems*, ACM, 9–15.
- [5] Dabbagh, N. & Kitsantas, A. 2012. Personal Learning Environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. *The Internet and Higher Education*. 15, 1, 3–8.
- [6] Dahlstrom, E., Walker, J.D., & Dziuban, C. 2013. *ECAR Study of Undergraduate Students and Information Technology*.
- [7] Downes, S. 2006. Learning Networks and Connective Knowledge. *Instructional Technology Forum*. Paper 92.
- [8] (ELI) Educause Learning Initiative. 2009. *Seven Things You Should Know About Personal Learning Environments*. Retrieved from: net.educause.edu/ir/library/pdf/ELI7049.pdf
- [9] Gillet, D. 2013. Personal learning environments as enablers for connectivist MOOCs. *12th Int. Conference on Information Technology Based Higher Education and Training*, 1–5.
- [10] Gruzdt, A. et al. 2014. Learning analytics for the social media age. *Proceedings of the 4th Int. Conference on Learning Analytics and Knowledge*. ACM, 254–256.
- [11] Gruzdt, A., & Haythornthwaite, C. 2011. Networking Online: Cyber communities. In *Scott, J., & Carrington, P. (Eds.), Handbook of Social Network Analysis*. London: Sage, 449–487
- [12] Haythornthwaite, C. 2001. Exploring Multiplexity: Social Network Structures in a Computer-Supported Distance Learning Class. *The Information Society*. 17, 3, 211–226.
- [13] Haythornthwaite, C. 2010. Multimodal, multi-actor literacies in e-learning environments. Paper presented at the *Multimodality and Learning Conference*, London, UK.
- [14] Ito, M., Gutierrez, K., Livingstone, S., et al. 2014. *Connected learning: An agenda for research and design*. Digital Media and Learning Research Hub, Irvine, CA.
- [15] Kop, R. et al. 2011. A Pedagogy of Abundance or a Pedagogy to Support Human Beings? Participant Support on Massive Open Online Courses. *International Review of Research in Open and Distance Learning*. 12, 7, 1–10.
- [16] Martindale, T. and Dowdy, M. 2010. Personal Learning Environments. *Emerging technologies in distance education*. G. Veletsianos, ed. Athabasca University Press. 177–193.
- [17] McLoughlin, C. and Lee, M.J.W. 2010. Personalised and self regulated learning in the Web 2.0 era: *International exemplars of innovative pedagogy using social software*. 26, 1, 28–43.
- [18] Moran, M., Seaman, J., & Tinti-Kane, H. 2012. *How Today's Higher Education Faculty Use Social Media*. Pearson/Babson Survey Research Group.
- [19] Paulin, D. and Haythornthwaite, C. (forthcoming). Crowdsourcing the Curriculum: Redefining E-Learning Practices Through Peer-Generated Approaches. (*Manuscript submitted for publication*).
- [20] Siemens, G. 2005. Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*. 2, 1, 3–10.
- [21] Siemens, G. 2008. Learning and knowing in networks: Changing roles for educators and designers. *ITFORUM for Discussion*, 1–26.
- [22] Smith, S.D. and Caruso, J.B. 2010. *The ECAR Study of Undergraduate Students*.
- [23] Vu, X.T. et al. 2014. Knowledge and Systems Engineering. 245, 109–119.
- [24] Wenger, E., Trayner, B., & de Laat, M. 2011. Promoting and assessing value creation in communities and networks: A conceptual framework. *The Netherlands: Ruud de Moor Centrum*.
- [25] Xu, J. et al. 2014. Quantifying cross-platform engagement through large-scale user alignment. *Proceedings of the ACM Conference on Web Science*, 281–282.
- [26] Zafarani, R. and Liu, H. 2009. Connecting Corresponding Identities across Communities. *ICWSM*, 354–357.
- [27] R. Zafarani and H. Liu. 2013. Connecting users across social media sites: A behavioral-modeling approach. In *Proceedings of the 19th ACM SIGKDD Int. Conference on Knowledge Discovery and Data Mining*, 41–49.
- [28] Haythornthwaite, C., Gruzdt, A., Paulin, D., Gilbert, S. and Esteve del Valle, M. 2015. *Social media use in teaching: Results from a questionnaire on use in higher education*. Social Media in Higher Education conference, Sheffield, UK.