How do you connect? Analysis of Social Capital Accumulation in connectivist MOOCs

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

Nia Dowell
Institute for Intelligent Systems
The University of Memphis
Memphis, USA
ndowell@memphis.edu

Oleksandra Skrypnyk
School of Education
University of South Australia
Adelaide, Australia
oleksandra.skrypnyk@mymail.unisa.edu.au

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

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

Shane Dawson
Learning and Teaching Unit
University of South Australia
Adelaide, Australia
shane.dawson@unisa.edu.au

Arthur C. Graesser
Department of Psychology
The University of Memphis
Memphis, USA
graesser@memphis.edu

ABSTRACT

Connections established between learners via interactions are seen as fundamental for connectivist pedagogy. Connections can also be viewed as learning outcomes, i.e. learners' social capital accumulated through distributed learning environments. We applied linear mixed effects modeling to investigate whether the social capital accumulation interpreted through learners' centrality to course interaction networks, is influenced by the language learners use to express and communicate in two connectivist MOOCs. Interactions were distributed across the three social media, namely Twitter, blog and Facebook. Results showed that learners in a cMOOC connect easier with the individuals who use a more informal, narrative style, but still maintain a deeper cohesive structure to their communication.

Categories and Subject Descriptors

Education; K.3.1 [Computer Uses in Education] Distance learning

General Terms

Social Processes, Automated Text Analysis, Learning

Keywords

Social capital, Language, Coh-Metrix, MOOCs, Social Network Analysis

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/10.1145/2723576.2723604

1. INTRODUCTION

Connectivist Massive Open Online Courses (cMOOCs) scale learner interactions by sharing, aggregating, and connecting information through the use of a diverse set of media. This approach allows learners to interact with each other around personal goals and common interests, outside of the teacher-controlled environment [1]. However, the distributed and open nature of cMOOCs complicates research inquiries into learning-related processes occurring in such environments. To date, the majority of cMOOC research has relied on self-reported mechanisms such as course evaluations obtained through participant surveys and identification of skills and capabilities that effectively support learner participation [2-5].

The establishment of social ties with other learners through interactions mediated by technology is viewed as integral to the learning process in cMOOCs [6, 7]. The quality of the relationships between the learners in a networked environment can be understood through the concept of social capital [8]. Essentially, a large amount of social capital reflects strong and productive relationships, based on the common interests and shared understanding among the participants [9, 10]. In this study, we further rely on the concept of social capital to describe the individual learning outcomes that result from the user interactions in cMOOCs using social media. Given that the social network analysis focuses on the relationships between individuals, rather than individuals and their properties [10], it is commonly used to assess the social capital and estimate the opportunities and limitations inherent to an individual actors' position in a social network [11]. For example, in an analysis of Twitter-based interactions within a cMOOC, Skrypnyk et al. [12] reported that an increase in the number and density of the communication acts resulted in an increased percentage of participants sharing the "power and control" over the information flow with the original course facilitators. This further means that very quickly after the course started, several course participants emerged as the most influential "factors" in knowledge sharing and brokering information. Consequently, individual positions in a learner network may indicate a degree of influence or faster access to more human and technological resources in the particular course.

The present study investigated the influence of learners' linguistic and discourse patterns, using an automated text analysis tool, on the accumulation of social capital. We analyzed the social networks extracted from the learner interaction within the three social media platforms (i.e., Twitter, Facebook, and blogs) that were used in two cMOOCs - namely, CCK11 and CCK12, as defined shortly. Specifically, linear mixed effects modeling assessed the association between the accumulation of social capital (determined through SNA) in a cMOOC and the linguistic and discourse features [13] used by learners in the content created and shared in social media.

2. THEORETICAL BACKGROUND

2.1 Social Capital

Individual positioning inside a social structure often yields material and symbolic resources that can be organized in a network of relationships of mutual acquaintance and recognition [11]. The invariable benefits of such networks are often called social capital. In educational research, social capital has helped to explain how the frequency and quantity of learner interactions and relationships is related to academic performance and drop out [14, 15]. There are links between academic work, social capital theory, and learning. For example, Gašević et al. [16] reported that learners' social capital is associated with the academic performance, whereas Kovanović et al. [17] connected social capital with the social presence in the communities of inquiry.

Social network analysis as a theory and method allows us to translate the abstract metaphor of social capital into an observable construct [11]. Specifically, measuring network structural properties, such as centrality, we are able to assess indicators describing the social capital accumulated by each individual in the network of learners. Besides focusing on the structural properties, SNA also provides information about the nature of the relationship and the strength of ties between learners. Therefore, in this paper, the semantic meaning and value of the interactions defines a tie between them, whereas measures of node centrality (i.e., degree, closeness, betweenness, and eigenvalue) are used to obtain a multi-dimensional measure of learners' social capital.

2.2 Language Use

Researchers typically rely on structural properties, as measured through SNA, when exploring online interactions. However, the interactions themselves customarily take place in natural language. From this view, language and discourse play a unique role in computer-mediated learning environments. It is the predominate channel used by learners to exchange thoughts and content. Moreover, the connection between discourse and social ties within a network is well established in numerous anthropological, sociological and sociolinguistic studies (for a review, [10]). Automated linguistics analysis methods are particularly well suited for handling the increasing scale of educational data. In line with this, linguistic analysis could provide rich contextual information to the behavioral patterns derived through SNA techniques. However, the combination of these two analytical methods is notably scarce in the literature, although there are exceptions [18]. The goal of this paper is to fill the large gap between semantic content analyses and SNA. Specifically, we adopted a theoretically grounded, computational linguistic analysis approach in combination with SNA to explore student interactions within an xMOOC. Psychological frameworks of discourse comprehension and learning have identified the representations, structures, strategies, and processes at multiple levels of discourse [19]–[21]. Five levels have frequently been offered in these frameworks: (1) words, (2) syntax, (3) the explicit textbase, (4) the situation model (sometimes called the mental model), and (5) the discourse genre and rhetorical structure (the type of discourse and its composition). We embrace this multilevel approach to language and discourse in the current paper.

3. RESEARCH QUESTIONS

Language and discourse is a channel of communication that is central to the information exchange within a network of learners, so the features of the communication channel are the main determinants of the social and cognitive processes that evolve in social networks [22]. The main premise of this study is that language and the quality of social ties established between learners in a cMOOC are mutually dependent and correlated. Building on this premise, we defined our research question as follows:

RQ: *How does the language used by learners in a cMOOC influence the accumulation of the social capital?*

4. METHODS

4.1 Data

This study examined the learner interaction occurring within blogs, Twitter and Facebook social media from the 2011 and 2012 editions of the Connectivism and Connective Knowledge (CCK) course. Both offerings were facilitated over a 12 week period. Live sessions were delivered using Elluminate, while course resources were delivered via gRSShopper. For the purpose of the automated data collection, we used gRSShopper as the source of links to blogs and copies of tweets. Facebook data from the course's open group were collected using Facebook API in order to retrieve communication between course participants. Finally, in order to support the analysis of content created in multiple languages, messages posted in languages other than English were translated using Microsoft Translation API (around 5% of the all messages).

The total number of active learners (N_{cck11}=997, N_{cck12}=429)[1] was higher during the CCK11 course, which was also reflected in the number of the posts created within the CCK11 (N_{post11}=5711, M=2.59, SD=4.47) and CCK12 (N_{post12}=2951, M=3.41, SD=9.06). However, despite a smaller cohort in the CCK12 course, the participants demonstrated a higher average activity (Facebook: $N_{post11f}$=1755, $N_{post12f}$=61; blogs: $N_{post11b}$=1473, $N_{post12b}$=624). Twitter-mediated communication sustained similar high levels of activity for both courses ($N_{post11t}$=2483, $N_{post12t}$=2266).

4.2 Analyses

4.2.1 Social Network Analysis

We constructed 72 undirected weighted graphs to represent interactions independently mediated by the three media (i.e., Twitter, blogs and Facebook) for each week of the two courses. *Twitter-based* social networks included all the authors and mentions as nodes of the network. The edges between two nodes were created if an author was tagged within the tweet. For example, if a course participant @L1 mentioned learners @L2 and @L3 in a tweet, then the course Twitter network would contain @L1, @L2, and @L3 with the following edges: @L1-@L2, and @L1-@L3. Social graphs from *Facebook* and *blog* communication included authors of the posts, i.e. blog owners or Facebook post initiators, as well as authors of comments to either of these. Specifically, if a learner A1 created a blog or Facebook post, and then learners B1 and C1 added comments to that post, the corresponding network would include nodes A1, B1, and C1 with the following edges: A1-B1, and A1-C1. All the weekly

1 Number of students for courses under study, represents the number of active students that participated in communication using three social media platforms analyzed.

social graphs extracted, included authors who posted and/or commented within the given week only.

The concept of node centrality is commonly used to assess the importance of an individual node within the network [23]. Therefore, the following well-established SNA measures [24] were calculated for each learner in all the network graphs: **degree centrality** (i.e., the number of edges a node has in a network), **eigenvalue centrality** (i.e., the measure of influence of a given node on other nodes), **closeness centrality** (i.e., the distance of an individual node in the network from all the other nodes), and **betweenness centrality** (i.e., the number of shortest paths between any two nodes that pass via a given node). For the analyses of the social network variables we used *igraph 0.7.1* [25], a comprehensive R software package for network analysis.

4.2.2 Linguistic Analysis

In order to conduct linguistic analysis, we parsed all the learner generated posts across the three media in weekly chunks. Specifically, all the posts produced by Learner 1 using Twitter as a media, during the first week of a course, were treated as a single unit. However, all the text produced by the same learner on Facebook within the same week in the same course was treated as another unit. Discourse analyses were conducted using Coh-Metrix computational linguistic facility [13], [26]. Coh-Metrix is, arguably, the most comprehensive automated textual assessment tool that allows for analysis of higher level features of language and discourse [13], [26]. In this study, we calculated the following five Coh-Metrix principal components, for the each unit of analysis: **narrativity** (i.e., the extent to which the text is in the narrative genre), **deep cohesion** (i.e., the extent to which the ideas in the text are cohesively connected), **referential cohesion** (i.e., the extent to which explicit words and ideas in the text are connected with each other as the text unfolds), **syntactic simplicity** (i.e., sentences with few words and simple, familiar syntactic structures), and **word concreteness** (i.e., the extent to which content words that are concrete and meaningful).

4.2.3 Statistical Analysis

All the variables (i.e., centrality measures and Coh-Metrix principal components) were measured at the individual level, and the data were structured in a way that learners were nested within a course. Therefore, we adopted a mixed-effects modeling approach, which is a recommended method for analyzing such datasets [27], allowing for more stringent examination of the effect of language on centrality by controlling for the variance associated with individual students and course differences. Four different linear mixed-effects models were constructed (i.e., *centrality models*), one for each of the four *dependent variables*: eigenvalue, degree, closeness, and betweenness. *Independent* fixed effect variables included five Coh-Metrix principal components. Moreover, media (i.e., Twitter, Facebook, and blogs), week, and post count were included as fixed effects. However, given the scope of this paper, those variables are not defined and elaborated. To address the impact of individual variance within a model, the course and learners within a course were treated as random effects.

The best mixed effects regression model was selected through the several steps. Besides the model with all the fixed effects included, *null models* with the random effects (*student within course, and course slope*), but no fixed effects were also constructed. A comparison of the null model with the *centrality* models determined whether language predicts social dynamics above and beyond the random effects. Intraclass Correlation Coefficient (ICC), [28], Akaike Information Criterion (AIC) and a likelihood ratio test [29], were used to decide on the best fitting and most parsimonious model. An effect size (R^2) was also estimated for each model as a goodness-of-fit measure denoting variance explained [30].

All the statistical analyses were conducted using R v.3.0.1 software for statistical analysis with package lme4, for fitting linear mixed-effects models [31]. Each of the hypotheses specify a specific direction in the effect, therefore one-tailed tests were used for significance testing with an alpha level of .05.

5. RESULTS

5.1 Degree Centrality

The results of the likelihood ratio test between the two models supported the conclusion that the *degree model* yielded a significantly better fit than the *null model*, $\chi^2(19) = 1506.5$, $p<.001$. Results of the linear mixed-effect analysis (Table 1. Centrality scores as a function of Coh-Metrix text characteristics) revealed a significant main effect for *Narrativity,* $F(1, 3042.7) = 4.13$ $p = .042$, *Referential Cohesion,* $F(1, 2806.4) = 27.32$, $p < .001$, *Deep Cohesion,* $F(1, 3034.1) = 4.22$, $p = .040$, *Syntax Ease,* $F(1, 3033.8) = 4.49$, $p = .032$. Specifically, individuals with a significantly lower degree centrality expressed themselves with a higher degree of referential cohesion and text simplicity. However, the learners with higher centrality scores had higher deep cohesion and narrativity.

Table 1. Centrality scores as a function of Coh-Metrix text characteristics

Measure	**Degree**		**Eigenvalue**	
	β	***SE***	**β**	***SE***
Narrativity	**0.03***	**0.03**	0.03	0.003
Word Concreteness	-0.006	0.01	-0.01	0.001
Referential Cohesion	**-0.07*****	**0.01**	**-0.06*****	**0.001**
Deep Cohesion	**0.04***	**0.02**	0.008	0.002
Syntax Simplicity	**-0.03***	**0.04**	-0.009	0.003
Measure	**Closeness**		**Betweenness**	
	β	***SE***	**β**	***SE***
Narrativity	-0.0009	0.0004	0.02	2.82
Word Concreteness	-0.02	0.0002	-0.02	1.25
Referential Cohesion	0.008	0.0002	**-0.04***	**1.30**
Deep Cohesion	0.01	0.0003	0.03	2.12
Syntax Simplicity	-0.004	0.0004	**-0.04***	**3.10**

Note: * $p < .05$; ** $p < .001$. Standard error (*SE*). *N*= 3066.

5.2 Eigenvalue Centrality

Similar to the degree model, the likelihood ratio test between the *null* model and the *eigenvalue model* revealed a significantly better fit of the model that accounted for variation of students within different courses ($\chi^2(19)= 681.62$, $p<.001$). The model (Table 1. Centrality scores as a function of Coh-Metrix text characteristics) showed a significant negative effect of *Referential Cohesion,* $F(1, 2667.4) = 13.33$, $p < .001$. Specifically, learners who exhibited lower scores for referential cohesion had higher eigenvector centrality values.

5.3 Betweenness and Closeness Centrality

We initially fit the same models with respect to degree and eigenvalue centrality to investigate how linguistic features of computer-mediated communicative utterances predict **betweenness** and **closeness centrality**. The models with all fixed and random effects resulted with better overall goodness-of-fit measures (AICc, R2, and ICC). However, further investigation of the results for the random effects showed the perfect negative correlation between random effects specified. This indicates that the model overfitted the data [32]. Therefore, we decided to discard models with random slope and continue analysis with the simpler models (i.e., student within a course as a random effect).

The *closeness model* did not reveal any significant linguistic properties and therefore is not further elaborated. In the case of the *betweenness model,* the likelihood ratio test with the *null model* indicated a better fit of the model that included fixed and random effects ($\chi^2(19)= 390.28$, $p<.001$). Reflecting on the solution for the fixed effects, we were able to identify a significant negative effect of *Referential Cohesion*, $F(1, 3026.6) = 4.19$, $p =$

.041 and *Syntactic Ease*, $F(1, 3042.3) = 5.04$, $p = .025$. The results show that course participants who tended to use *simple linguistic constructs* with higher *referential cohesion* had lower betweenness centrality.

6. DISCUSSION

6.1 Interpretation of the results

The results indicate that deep level linguistic characteristics (i.e., Coh-Metrix indices) influence learner interaction within a cMOOC. This paper did not examine surface level features (e.g., count of posts), however it supports the claim that a systematic and deeper analysis (beyond the surface level dialogue characteristics) is necessary in order to obtain a more comprehensive insight into the linguistic processes that shape learning in network settings and influence the development of social connections [33].

The results suggested that linguistic and discourse features of written artefacts are important determinants of learning in a cMOOC environment. Specifically, our results show that learners whose discourse was more narrative, with deeper cohesion, more complex linguistic structures, and low referential cohesion had more connections, and interacted more often with their peer learners and instructors. Likewise, learners who authored posts with low values of referential cohesion had more ties with the most influential, well-connected learners, as indicated with higher values of eigenvalue centrality. Finally, higher potential for the control of communication and brokerage of information (i.e., higher betweenness centrality) included learners who tend to integrate new information (i.e., lower cohesion) within each post and who had a discourse that was more syntactically embedded.

Thus, what is the overall effect of deep level linguistic and discourse properties on the accumulation of learners' social capital in a cMOOC? Course participants who tend to use more narrative and informal style, nevertheless still manage to maintain a deeper cohesive structure in their communication will have more ties. That being said, we were able to conclude that *language does define structural positions* within the social network emerging from the interaction in network learning environment. The way the learners convey the messages and share the information, could potentially bring them benefits in terms of strengthening ties with peer learners and consequently increase the social capital [11].

It is also indicative that individuals who created posts with higher referential cohesion, attracted "attention" (i.e., comments and reactions) from fewer participants. Given that referential cohesion captures the extent to which ideas in the text are connected with each other as the text unfolds, higher referential cohesion indicates fewer gaps in conveying the ideas and increased text readability and comprehension [19]. On the other hand, referential cohesion gaps occur when a sentence has few if any words that overlap with previous sentences. [19]. A possible explanation for the relation of lower referential cohesion and possibly complex syntactic structure with increased social capital could be connected to the affordances of media used in the analyzed cMOOC, i.e. Twitter and Facebook posts were noticeably shorter than blog posts. This further implies that, in terms of overlap between sentences and paragraphs, paragraph-to-paragraph measures should be interpreted as post-to-post referential cohesion. In this context, the lower referential cohesion might be capturing a lack of overlap between an individual learner's posts. In this case, learners who tend to post more topically diverse messages would naturally have less overlap and consequentially lower referential cohesion values of their posts compared to their more topically uniform counterparts. Therefore, it is likely that learners who tended to provide novel information across their posts, attracted more peers and attained more "followers". Likewise, low referential cohesion across the discourse might indicate that those learners triggered many discussions about dilemma's and challenging topics. High referential cohesion might indicate the redundant information across the posts with lack of the "real" contribution to discussions and knowledge development. Such interpretations should, however, be taken with a degree of caution until further studies test replications and relevant interpretations.

6.2 Implications for theory and practice

Our findings have shown that learners' ability to effectively use language to communicate and share knowledge with peers is essential to the building new ties and strengthening existing connections. Moreover, being able to recognize important information and coherently develop new ideas building on the existing knowledge ultimately leads to the accumulation of social capital. Finally, observed from the linguistic perspective, sharing novel information, using concrete and coherently structured language (i.e., written text) is perhaps the main prerequisite for establishing new connections within the network of learners. It seems that in highly distributed environments of cMOOCs, learners tend to value new information, new ideas, triggering novel potentially interesting and relevant discussions, rather than elaborating on a single topic (or a small number of topics) throughout a course. However, these new information have to be comprehensive, well-structured in order to increase understanding among learners and foster the interaction. Nevertheless, further research is needed to assess individuals' ability not only to develop a social capital, but also to take advantages of the accumulated social capital for a specific return (e.g., to facilitate the achievement of learning outcomes).

It is questionable whether learners would be able to develop all the necessary skills for learning in networked environment simply by interacting with their peers. Therefore, future research needs to investigate various instructional scaffolds and available technological affordances that would provide guidelines for students in developing necessary skills for learning in such settings. Those skills, identified as "new media literacies" [34], should enable learners to use media affordances more efficiently thus gaining more from learning in distributed learning contexts. Eventually, changes in the way learners use the linguistic features could provide an insight into individual's progress in the development of those literacies. On the other hand, as indicated by various studies on online and distance education, personalized, formative and timely feedback presents one of the most promising approaches for fostering learning in online settings [35, 36]. Information gleaned from these findings suggests discourse analytics could prove useful in creating personalized feedback for students interacting within computer-mediated, networked platforms in the future. For instance, a system could provide accurate real time support for learners using an interface that delivered suggestions via a simple pop up window or a more sophisticated intelligent agent. Such computer-mediated support could help course participants develop improved information transfer and gathering skills. However, the value of such enhancements awaits future work and empirical testing. Nevertheless, this research might represents an initial step, by highlighting potentially useful analytical tools, and stimulating discussion about MOOC platforms capable of enhanced dynamic social processing, and automated cognitive evaluation for learner feedback.

6.3 Limitation

It is important to acknowledge limitations of this study. At the time we were collecting the data for the analysis (April-August 2014), tweets posted within the both courses under study were no longer available through Twitter API. Given that we obtained those data using gRSShopper as a source, some of the course interactions, such as replies, retweets**,** and favorites, could not be collected. However, features such as mentions and hashtags, along with the tweet content, were preserved. Additionally, the study analyzed the data from a course in a specific subject domain. It is reasonable to assume that different subject domains would be characterized with different communication patterns. Therefore, it would be prudent to analyze social interactions within courses from various subject domains.

7. CONCLUSIONS

Through deep levels of text analyses, our findings show that linguistic and discourse features have a significant impact on the accumulation of learners' social capital in a networked learning setting. The findings suggest that facilitators of distributed courses should consider a broad array of responsibilities that include and extend simple network-formation beyond shaping and leveraging the information flows throughout the learning network. For example, cMOOC facilitators could introduce instructional elements that enhance group and individual communication skills. Finally, the study opens up for further investigation of the relationship between social ties and language in use.

8. REFERENCES

[1] A. A. Mcauley, B. Stewart, G. Siemens, and D. Cormier, "The MOOC Model for Digital Practice," p. 33, 2010.

[2] 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.

[3] 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.

[4] R. Kop, J. Sui, and 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. 1–10, 2011.

[5] C. Milligan, A. Littlejohn, and A. Margaryan, "Patterns of Engagement in a Connectivist MOOC," *J. Online Learn. Teach.*, vol. 9, no. 2, 2013.

[6] J. Anh, B. Butler, and A. Alam, "Learner Participation and Engagement in Open Online Courses: Insights from the Peer 2 Peer University," *MERLOT JOLT*, vol. 9, no. 2, 2013.

[7] J. Knox, "Digital culture clash: 'massive' education in the E-learning and Digital Cultures MOOC," *Distance Educ.*, vol. 35, no. 2, pp. 164–177, 2014.

[8] N. Lin, "Social capital: a theory of social structure and action," *Soc. Forces*, vol. 82, no. 3, pp. 1209–11, 2004.

[9] A. Lockwood, "Community collaboration and social capital: an interview with Gary G. Wehlage," *Lead. Tomorrows Sch.*, vol. 2, pp. 19–25, 1996.

[10] J. Scott and P. Carrington, *The SAGE Handbook of Social Network Analysis*. SAGE, 2011.

[11] R. S. Burt, "The Network Structure of Social Capital," *Res. Organ. Behav.*, vol. 22, pp. 345–423, 2000.

[12] O. Skrypnyk, S. Joksimovic, V. Kovanovic, D. Gasevic, 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.

[13] A. Graesser, D. Mcnamara, and J. Kulikowich, "Coh-Metrix: Providing Multilevel Analyses of Text Characteristics," *Educ. Res.*, vol. 40, no. 5, pp. 223–234, Jun. 2011.

[14] C. Carceller, S. Dawson, and L. Lockyer, "Improving academic outcomes: does participating in online discussion forums payoff?," *Int. J. Technol. Enhanc. Learn.*, vol. 5, no. 2, pp. 117–132, 2013.

[15] L. M. Vaquero and M. Cebrian, "The rich club phenomenon in the classroom," *Sci Rep*, vol. 3, Jan. 2013.

[16] D. Gašević, A. Zouaq, and R. Janzen, "'Choose Your Classmates, Your GPA Is at Stake!': The Association of Cross-Class Social Ties and Academic Performance," *Am. Behav. Sci.*, 2013.

[17] V. Kovanović, S. Joksimović, D. Gašević, and M. Hatala, "What is the Source of Social Capital? The Association between Social Network Position and Social Presence in Communities of Inquiry," in *Proceedings of the Workshops held at Educational Data Mining 2014, co-located with 7th International Conference on Educational Data Mining (EDM 2014)*, London, UK, 2014, pp. 1–8.

[18] A. J. Scholand, Y. R. Tausczik, and J. W. Pennebaker, "Assessing Group Interaction with Social Language Network Analysis," in *Advances in Social Computing*, S.-K. Chai, J. J. Salerno, and P. L. Mabry, Eds. Springer Berlin Heidelberg, 2010, pp. 248–255.

[19] A. C. Graesser and D. S. McNamara, "Computational Analyses of Multilevel Discourse Comprehension," *Top. Cogn. Sci.*, vol. 3, no. 2, pp. 371–398, 2011.

[20] W. Kintsch, *Comprehension: A Paradigm for Cognition*. Cambridge, U.K.: Cambridge University Press, 1998.

[21] C. E. Snow, *Reading for Understanding: Toward a Research and Development Program in Reading Comprehension*. Santa Monica, CA: Rand Corporation, 2002.

[22] N. Dowell, W. Cade, Y. Tausczik, J. Pennebaker, and A. Graesser, "What Works: Creating Adaptive and Intelligent Systems for Collaborative Learning Support," *Intell. Tutoring Syst. Proc. ITS 2014*, pp. 124–133, 2014.

[23] L. C. Freeman, "Centrality in social networks conceptual clarification," *Soc. Netw.*, vol. 1, no. 3, pp. 215–239, 1979.

[24] S. Wasserman and K. Faust, *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press, 1994, p. 825.

[25] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *InterJournalournal*, vol. Complex Sy, p. 1695, 2006.

[26] D. S. McNamara, A. C. Graesser, P. M. McCarthy, and Z. Cai, *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, M.A.: Cambridge University Press., 2014.

[27] J. C. Pinheiro and D. M. Bates, *Mixed-effects models in S and S-PLUS*. Springer, 2000.

[28] S. W. Raudenbush and A. S. Bryk, *Hierarchical Linear Models: Applications and Data Analysis Methods*. SAGE Publications, 2002.

[29] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd editio. New York: Springer, 2009.

[30] R. Xu, "Measuring explained variation in linear mixed effects models," *Stat. Med.*, vol. 22, no. 22, pp. 3527–3541, 2003.

[31] D. Bates, M. Maechler, B. Bolker, and S. Walker, *lme4: Linear mixed-effects models using Eigen and S4*. 2013.

[32] R. H. Baayen, *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge University Press, 2008.

[33] M. De Laat, "Networked learning," *Police Acad. Neth. Apeldoorn*, 2006.

[34] S. Dawson and G. Siemens, "Analytics to literacies: The development of a learning analytics framework for multiliteracies assessment," *Int. Rev. Res. Open Distance Learn.*, vol. 15, no. 4, 2014.

[35] R. M. Bernard, P. C. Abrami, E. Borokhovski, C. A. Wade, R. M. Tamim, M. A. Surkes, and E. C. Bethel, "A Meta-Analysis of Three Types of Interaction Treatments in Distance Education," *Rev. Educ. Res.*, vol. 79, no. 3, pp. 1243–1289, Sep. 2009.

[36] E. Borokhovski, R. Tamim, R. M. Bernard, P. C. Abrami, and A. Sokolovskaya, "Are contextual and designed student–student interaction treatments equally effective in distance education?," *Distance Educ.*, vol. 33, no. 3, pp. 311–329, 2012.