Online Knowledge Communities as Student-Centered Open Learning Environments: How Likely Will They Be to Integrate Learners as New Members?

Nicolae Nistor, Universität der Bundeswehr München, Germany/Walden University, USA, nic.nistor@lmu.de
Mihai Dascălu, University Politehnica of Bucharest, Romania, mihai.dascalu@cs.pub.ro
Christian Tarnai, Universität der Bundeswehr München, Germany, christian.tarnai@unibw.de
Nicolas Bresser, Universität der Bundeswehr München, Germany, nicolas.bresser@gmx.de
Ştefan Trăuşan-Matu, University Politehnica of Bucharest, Romania, stefan.trausan@cs.pub.ro

Abstract: Using online knowledge communities (OKCs) from the Internet as student-centered, open learning environments (SCOLEs) poses the question how likely these communities will be to integrate learners as new members. This premise of learning in SCOLEs is analyzed in the current study. Based on the approaches of voices interanimation and polyphony, a natural language processing tool was employed for dialog analysis in integrative vs. non-integrative blog-based OKCs. Three dialog dimensions were identified: participants' individual content-oriented contribution, social contribution, and their position within the social network. Hierarchical clusters built upon these dimensions reflect socio-cognitive structures including central, regular and peripheral OKC members. OKCs with a stronger layer of regular members appear more likely to integrate new members, whereas OKCs with a stronger layer of peripheral members appear less likely to do so. Consequently, the study suggests an automated prediction method of OKC integrativity that may sustain the educational use of OKCs.

Keywords: virtual communities of practice, student-centered open learning environments (SCOLEs), natural language processing, dialogue analysis, polyphony, interanimation

Introduction

Online knowledge communities (OKCs) are frequently regarded in educational research as collaborative learning environments, for instance in the particular shape of student-centered, open learning environments (SCOLEs; Hannafin, Hill, Land, & Lee, 2014). However, using existing OKCs from the Internet as SCOLEs poses the question how likely these communities will be to integrate learners as new, legitimate peripheral members (Lave & Wenger, 1991). The following study contributes to understanding and predicting this premise of learning in SCOLEs, which will further provide a tool and method to sustain the educational use of OKCs.

Theoretical background

SCOLEs are a special case of learning environments in which the individual learner determines his or her learning goals, learning means, or both. The learner is thus central in defining meaning; participation is scaffolded in authentic, often ill-structured tasks and socio-cultural practices; learners access diverse resources, roles and perspectives; prior individual experiences are the main ground of knowledge construction (Hannafin et al., 2014). Nistor et al. (in press) posit that SCOLEs can be designed as OKCs, in which students may participate in the community discourse and thus collaboratively learn while applying domain-specific knowledge and argumentation strategies. As described by several authors (Bereiter & Scardamalia, 2014), knowledge building communities display multi-layered socio-cognitive structures. Accordingly, Lave and Wenger (1991) make a difference between (1) central, (2) regular and (3) peripheral participation. Accordingly, central participants assume more responsibility and perform more difficult tasks than peripheral members; therefore their identity is that of an expert. A more recent classification with finer granularity (Wenger-Trayner, 2011) indicates five categories: (1a) leaders and (1b) core group, regarded as sub-categories of the central participants' group; (2) active or regular participants; and finally (3a) occasional participants and (3b) transactional participants, two relatively distinct groups at the periphery of the community of practice (CoP).

After a few decades of mainly qualitative research on communities, researchers are beginning to apply quantitative methods. This line of research is mainly illustrated by Borgatti et al. (2009) who apply social network analysis (SNA) to describe group social structure. Nistor and Fischer (2012) apply SNA in combination with further quantitative methods to validate a model of the relationship between expertise, participation and expert status in academic communities of practice. In the newly emerged domain of Learning Analytics, particularly in Discourse Analytics, methods including SNA, clustering, and factor analysis are applied to

identify socio-cognitive structures and predict learning in technology-based learning environments (Baker & Siemens, in press). In this vein, Nistor et al. (in press) use an automated dialog analysis tool, *ReaderBench* (Dascălu et al, 2013; Dascălu, 2014), to identify and assess OKCs as potential SCOLEs on the Internet.

The automated assessment of OKCs is based on the idea that community discourse is tightly connected with socio-cognitive structures, practice and learning (Hannafin et al., 2014; Lave & Wenger, 1991). The *ReaderBench* tool is based on Bakhtin's (1981) dialogism and on the polyphonic model of discourse (Trău□an-Matu, 2010). The tool is described in detail by Dascălu (2014) and validated by Nistor et al. (in press). It provides several indicators describing the personal and social dimensions of a collaborative dialog, emphasizing dialogue coherence and overall coverage of a given topic. These dimensions are strongly correlated with participants' expertise (Nistor et al., 2014) and critical thinking (Nistor et al., in press) expressed in online, text-based discussions.

Such methods and tools evidently support the educational use of OKCs as collaborative learning environments. However, a closer examination of the idea to use existing OKCs from the Internet as SCOLEs suggests at least two design possibilities for the educational practice. In one of them, OKCs may be created for educational purposes in line with a given curriculum. In the other, learners may peripherally participate (Lave & Wenger, 1991) in existing OKCs from the Internet. Given that communities need relatively long time to form and become functional, the former option may be very restrictive if only short time (e.g., one term or academic year) is available. The latter option seems less restrictive from this point of view, however, it raises a major question: *How likely is an existing OKC to integrate learners as new, legitimate peripheral members?* The following study aims to answer this research question.

Methodology

The study explores the socio-cognitive structures of OKCs that were likely vs. unlikely to integrate newcomers (in the following called integrative vs. non-integrative OKCs), following three steps: (1) analyze the community discourse using the *ReaderBench* tool (Dascălu et al, 2013; Dascălu, 2014); (2) cluster the community members based on the resulting discourse characteristics; (3) compare the clustering results in integrative vs. non-integrative virtual communities.

The analysis was conducted on the Internet, in blogger communities publicly available on the blogspot.com and wordpress.com platforms. In a prior study, the researchers had attempted to initiate discussions in several blog communities. Two situations emerged: one in which the blog participants responded to the initiated dialogue, another in which the communication attempt was ignored. Consequently, it was assumed that the former group consisted of integrative (n = 5), the latter of non-integrative (n = 4) OKCs. After these N = 9 blogger communities with a total of 308 participants were chosen, the entire community discourse produced within three months (beginning with the researchers' communication attempt) was downloaded and automatically analyzed. No personal data of the participants was collected.

The *ReaderBench* tool provides 13 dialog indicators: two overall indicators (Number of comments, Total collaborative dialogue quality), one indicator of the individual contribution to the dialog (Individual collaborative dialogue quality), five indicators of the social contribution to the dialog (Number of initiated threads, Length of initiated threads, Cumulative interanimation of voices, Social collaborative dialogue quality, Social collaborative dialogue quality in initiated threads), and five SNA centrality indicators (Indegree, Outdegree, Closeness, Eccentricity, Betweenness).

Findings

Discourse analysis

The absolute values of the variables ranged in large limits; hence they were standardized. Further on, they were strongly correlated with each other; therefore, a principal component analysis was performed. Thus, the number of components was reduced to three factors with eigenvalues greater than 1, which explains 80.35% of the total variance. The three dimensions resulting after varimax rotation are based on different sets of the initial variables, as follows. *Factor 1R* is mainly based on the Quality and Number of Initiated Discussions, the Individual Collaborative Dialogue Quality, Outdegree and Betweenness. As such, Factor 1R is related to the individual *content-oriented contribution* to the dialogue. *Factor 2R* is mainly composed of the Social Collaborative Dialogue Quality, the Cumulative Interanimation, Indegree, and the number of comments. Thus, it describes the individual *social contribution* to the collaborative dialogue. *Factor 3R* mainly includes the SNA variables Closeness and Eccentricity, therefore it describes the individual *position within the social network*.

Cluster analysis

In the second step of the analysis, the three dimensions resulting from the principal component analysis (Anderson-Rubin method) were used as input for a hierarchical cluster analysis according to the Ward method with quadratic Euclidian distances. The optimal separation of clusters was reached for the following five clusters.

First, Clusters 4 and 1 are most visible due to the high individual contribution to the dialogue (Factor 1R). Cluster 4 consists of n = 3 participants, who are all blog owners and, as such, initiate the most and the longest discussions. Cluster 1 consists of n = 13 participants, including 6 blog co-owners from blog 5, the only analyzed blog that has multiple owners. These participants are highly active with respect to initiating and sustaining discussions, however somewhat less than Cluster 4, nevertheless more than all other clusters. For these reasons, Cluster 4 and 1 reunite the *central OKC members*, from which Cluster 4 includes OKC leaders, and Cluster 1 the core group of the OKC.

Second, Cluster 2 includes n = 42 blog members with the strongest contribution to the social collaborative quality of the dialogue (Factor 2R). They even contribute more to the interanimation of voices, and the display higher Indegree than Cluster 1. These appear to be the *regular* or *active OKC members*.

Third, two large clusters, Cluster 3 (n = 157) and Cluster 5 (n = 93) reunite the least active OKC members. The blog participants from Cluster 5 are positioned at largest distance from the OKC center, they never initiate discussions, and they participate only occasionally in existing discussions; however, they contribute new information to the discussions. In contrast, Cluster 3 participants contribute to the interanimation of voices within the dialogue. These, as well, exercise little activity and stand at the periphery of the social network, however not as far from the center as Cluster 5. Hence, Cluster 3 and Cluster 5 can be described as *peripheral OKC members*. From these, Cluster 3 seems to correspond to occasional participants, and Cluster 5 to transactional participants.

Integrative versus non-integrative blogger communities

By comparing the identified clusters between integrative and non-integrative blog communities, it appears that integrative OKCs include a larger active subgroup of regular members than non-integrative OKCs. In 4 out of 5 integrative OKCs the regular:peripheral proportion is higher than 10%, whereas in 3 out of 4 non-integrative OKCs the same proportion lays under 10%. Choosing a threshold value of 10% leads to a correct integrative vs. non-integrative classification for 7 out of 9 blogs. Blog 3 becomes then a false non-integrative, Blog 9 a false integrative blog.

Discussion

In summary, this study lays the ground for the design and development of OKCs as SCOLEs (Hannafin et al., 2014), in which learners are supposed to go out of traditional or virtual classrooms, into existing OKCs, discuss topics of interest, come back with results, and draw conclusions. This requires that the OKCs integrate the learners in their community discourse. This study assumes that the integrativity of an OKC is tightly connected to the community discourse and practice, hence it can be assessed by discourse analysis, as follows.

In the first step, the polyphony-based (Trău□an-Matu, 2010) tool *ReaderBench* (Dascălu et al, 2013; Dascălu, 2014) was employed to analyze the blog-based OKC discourse. From the multitude of provided results, the following ground dimensions were extracted: (1) individual participants' content-oriented contribution, (2) their social contribution to the collaborative dialogue, and (3) their position within the social network. These dimensions result from Bakhtin's (1981) polyphony theory and Trău□an-Matu's (2010) analytic approach. They describe the interanimation of voices within a collaborative dialogue, and appear appropriate for automated discourse analysis.

In the second step, the community members were clustered based on their discourse characteristics (Nistor et al., 2014). The hierarchical cluster analysis offered two classifications that both correspond to the socio-cognitive structures described in the CoP research. A classification with lower granularity reflects the same community layers described in the early CoP literature (Lave & Wenger, 1991): (1) central, (2) regular/active and (3) peripheral members. A classification with higher granularity indicates five categories that correspond to later CoP literature (Wenger-Trayner, 2011): (1a) leaders and (1b) core group, that can be regarded as sub-categories of the central participants' group, again the – relatively homogeneous – category of (2) regular/active participants, and the categories of (3a) occasional and (3b) participants, as two relatively distinct groups at the OKC periphery.

In the third and final step, knowing from prior experiments the integrativity of the analyzed OKCs, the clustering results were compared and the decision threshold (integrative vs. non-integrative) was determined. This threshold indicates that a minimum number of regular members (at least ten percent of the periphery) is

necessary for the OKC to be integrative. Thus, the role and importance of the regular/active members becomes evident. In line with Lave and Wenger's (1991) observations, it is this community subgroup that mainly carries out the daily tasks and activities, including the integration of new members. Notably, no significant differences in the (very small) central member groups were observed.

Conclusions

For the practice of computer-supported collaborative learning, the conclusions of this study are straightforward: Existing OKCs from the Internet can be used as SCOLEs, for example in higher education. Appropriate instructional design should be developed and evaluated in the near future. For OKC research, this study adds empirical evidence for the existence and importance of finer socio-cognitive structures, and for the relationship between community discourse and practice. However, the result validity may be limited by several conceptual and methodological shortcomings. Although the number of participants was fairly high, there were a small number of OKCs involved in the study. Also, integrativity was assimilated to OKC members' response to relatively simple information requests. Upcoming research aims to extend both the samples and the perspective on integrativity by observing the long-term interactions between regular OKC members and visitors.

References

Baker, R., & Siemens, G. (in press). Educational data mining and learning analytics. *Cambridge handbook of the learning sciences* (2nd edition). Cambridge, UK: Cambridge University Press. http://www.columbia.edu/~rsb2162/BakerSiemensHandbook2013.pdf

Bakhtin, M. M. (1981). *The dialogic imagination: Four essays*. London: The University of Texas Press.

Bereiter, C. & Scardamalia, M. (2014). Knowledge building and knowledge creation: One concept, two hills to climb. In S. C. Tan, H. J. So, & J. Yeo (Eds.), *Knowledge creation in education* (pp. 35-52). New York: Springer.

Borgatti, S. P., Mehra, A., Brass, D. J. & Labianca, G. (2009). Network analysis in the social sciences. *Science, 323*, 892-895.

Dascălu, M. (2014). *Analyzing discourse and text complexity for learning and collaborating. Studies in Computational Intelligence* (Vol. 534). Switzerland: Springer.

Dascalu, M., Trausan-Matu, S., & Dessus, P. (2013). Cohesion-based analysis of CSCL conversations: Holistic and individual perspectives. In N. Rummel, M. Kapur, M. Nathan, & S. Puntambekar (Eds.), *10th International Conference on Computer-Supported Collaborative Learning 2013* (pp. 145–152). Madison, USA: International Society of the Learning Sciences.

Hannafin, M. J., Hill, J. R., Land, S. M., & Lee, E. (2014). Student-centered, open learning environments: Research, theory, and practice. In J. M. Spector, M. D. Merrill, J. Elen & M. J. Bishop (Eds.). *Handbook of Research on Educational Communications and Technology* (pp. 641–651). New York: Springer.

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

Nistor, N., Baltes, B., Smeaton, G., Dascălu, M., Mihăilă, D., & Trău□an-Matu, □. (2014). Participation in virtual academic communities of practice under the influence of technology acceptance and community factors. A learning analytics application. *Computers in Human Behavior, 34*, 339–344.

Nistor, N. & Fischer, F. (2012). Communities of practice in academia: Testing a quantitative model. *Learning, Culture and Social Interaction, 1*(2), 114–126.

Nistor, N., Trăuşan-Matu, Ş., Dascălu, M., Duttweiler, H., Chiru, C., Baltes, B., & Smeaton, G. (in press). Finding student-centered open learning environments on the internet: Automated dialogue assessment in academic virtual communities of practice. *Computers in Human Behavior*. DOI: 10.1016/j.chb.2014.07.029

Trău□an-Matu, □. (2010). The polyphonic model of hybrid and collaborative learning. In F. Wang, L. J. Fong, & R. C. Kwan (Eds.), *Handbook of research on hybrid learning models: Advanced tools, technologies, and applications* (pp. 466–486). Hershey. NY: Information Science Publishing.

Wenger, E. (1998). *Communities of practice. Learning, meaning, and identity*. Cambridge, UK: Cambridge University Press.

Wenger-Trayner, E. (2011). *Levels of participation*. http://wenger-trayner.com/wp-content/uploads/2011/12/11-12-Levels-of-participation-600.jpg

Acknowledgments

This work has been partially funded by the Sectorial Operational Program Human Resources Development 2007-2013 of the Romanian Ministry of European Funds according to the Financial Agreement POSDRU/159/1.5/S/134398.