

Informal Learning in Online Knowledge Communities: Predicting Community Response to Visitor Inquiries

Nicolae Nistor^{1,2,3}, Mihai Dascălu⁴, Lucia Larise Stavarache⁴, Yvonne Serafin²,
and Ștefan Trăușan-Matu⁴

¹ Faculty of Psychology and Educational Sciences, Ludwig-Maximilians-Universität,
Leopoldstr. 13, 80802 Munich, Germany

nic.nistor@uni-muenchen.de

² Faculty of Human Sciences, Universität der Bundeswehr München,
Werner-Heisenberg-Weg. 39, 85577 Neubiberg, Germany

yvonne.serafin@unibw.de

³ Richard W. Riley College of Education and Leadership Walden University,
100 Washington Avenue South, Suite 900, Minneapolis, MN 55401, USA

⁴ Faculty of Automatic Control and Computers, University "Politehnica" Bucharest,
Splaiul Independenței 313, 60042 Bucharest, Romania

{mihai.dascalu, stefan.trausan}@cs.pub.ro,

larise.stavarache@ro.ibm.com

Abstract. Informal learning in online knowledge communities (OKCs) comprises visitor inquiries on specific topics. Learning can occur only if the OKC adequately respond. This study aims to predict OKC response, using a social learning analytics approach based on computational linguistics and Bakhtin's theory of dialogism. Observing the blog topic (cooking vs. politics & economics) and the visitor inquiry format (off-topic vs. on-topic), a field experiment with a 2×2 factorial design was conducted on a sample of $N = 68$ blogger communities with a total of 25,303 members. For the entire sample, the community response was influenced only by the inquiry format. In a separate examination of experimental groups, only for one examined topic (cooking) this remained true, while for the other (politics & economics) the community response only depended on the previously established dialog quality. The findings suggest identification criteria for responsive communities, which can support OKC integration in learning environments.

Keywords: Social learning analytics · Computational linguistics · Informal learning · Online knowledge communities

1 Introduction

People search and visit online knowledge communities (OKC) to ask questions, and clarify problems encountered in their daily life. From an educational perspective, this can be regarded as informal learning. However, it cannot be taken for granted that an OKC will adequately respond. Successful informal learning in OKCs would strongly benefit from OKC responsiveness. Social learning analytics can provide an automated

means to predict OKC responsiveness [1]. A first step towards this endeavor was taken by Nistor, Dascălu, Tarnai, Bresser and Trăușan-Matu [2]. This paper proposes a refinement and empirical validation of their approach.

2 Theoretical Framework

OKCs are groups of mutually engaged people, sharing interests, knowledge and “ways of doing things”, and sustaining a joint computer-mediated discourse over longer periods of time [3, 4]. A central aspect of OKCs resides in the socio-cognitive structure that includes central, active and peripheral members [4]. Central participants assume more responsibility and perform more difficult tasks than peripheral participants; therefore, their identity is that of an expert. Identity within the community typically develops from newcomer and novice to old-timer and expert, and is negotiated in the community discourse (*ibid.*). The inseparability of knowledge, identity and social practice on the one hand, and community discourse on the other (*ibid.*), suggests that adequate discourse analysis may inform the researcher about the socio-cognitive structures of OKCs, and predict the probability of new member integration.

In this vein, Nistor and colleagues [2] conducted an automated discourse analysis based on Bakhtin’s dialogism [5] and on the polyphonic discourse model [6, 7]. Measuring participants’ involvement has been computed as follows. Firstly, the relevance of each contribution is identified using the aggregate score of each constituent word. Each word score considers three analysis dimensions [8]: (1) statistical presence, (2) semantic relatedness, and (3) semantic coverage derived from the importance of the semantic chain covering the given concept and spanning throughout the conversation thread. Secondly, each post or comment is assigned an individual score equal to the normalized term frequency of each word multiplied by its relevance [9]. Finally, the overall contribution score (either for the initial post or of the corresponding comments) is computed according to its on-topic relevance. Additionally, each contribution score is augmented through its connections or cohesive links to other posts or comments, thus emphasizing the information transfer between utterances [9].

Transposing the local importance of each concept with respect to other inter-linked utterances models the inter-animation of ‘voices’ or opinions [10]. Furthermore, in order to measure the participation engagement in the discussion thread, the voices of different participants have been overlapped in a polyphonic manner [11]. Thus, by applying a pointwise mutual information (PMI) model on the moving averages of voice distributions [9] from different participants, a reliable estimator of the degree of inter-animation between voices results [10].

Altogether, the resulting score per community member is computed as the aggregated of the above from each individual conversation thread where the members contributed, thus generating a global overview of member participation in an OKC. The automated computational measurements have been implemented in the Reader-Bench tool, described in detail by Dascălu [9] and validated by Nistor et al. [2]. Several indicators describing the personal and social quality dimensions of a collaborative dialog, emphasizing discourse cohesion and the overall coverage of given topics have been used in order to compute the scores.

Employing this discourse analysis tool, Nistor and colleagues [2] propose a prediction criterion of OKC responsiveness and integrativity; however, their study has some limitations related to its quasi-experimental approach, the operationalization of community response, and sample size. Further research is needed to overcome these limitations.

3 Research Questions and Methodology

The research addressed above suggests that the OKC response to visitor inquiries may depend on the visitor inquiry format, in addition to the socio-cognitive OKC structure, and the collaborative dialog quality previously established in the community. This assumption is explored through the following research questions:

- RQ1. What is the effect of the socio-cognitive structure on the collaborative dialog quality in blogger communities?
- RQ2. What is the effect of visitor inquiry format, blog topic, collaborative dialog quality, and socio-cognitive structure of blogger communities on the community response?

Accordingly, a field experiment with a 2×2 factorial design was conducted within blogger communities on the Internet. One of the factors was the treatment, i.e., the *visitor inquiry format*. This consisted of either an off-topic inquiry (a request for participation in a questionnaire survey); or an on-topic inquiry consisting of a short dialog (three question-answer or answer-comment sequences), followed by the same request for participation in a questionnaire survey. The second factor was the *blog topic*. Two different topics were chosen: cooking, and politics & economics, with the expectation that participants would employ different communication, argumentation and knowledge construction strategies for each topic [12].

A sample of $N = 68$ blogger communities was analyzed, for a total of 25,303 members distributed between 1 and 1844 per community (in average $M = 7.54$, $SD = 9.55$ participants per month of blog lifetime). The blogs were distributed across experimental groups as shown in Table 1.

Table 1. Distribution of blogs across the 2×2 factorial design.

Blog topics	Treatment 1: off-topic inquiry	Treatment 2: on-topic inquiry
Cooking	20	17
Politics & economics	16	15

Besides treatment and blog topic, several covariates were further considered. Seven indicators of the *quality of collaborative dialog* at the blog level were provided by the ReaderBench tool [9, 13]. From these indicators, the *socio-cognitive structure* of the examined OKCs, i.e., the number of central, active and peripheral members, was subsequently

extracted according to the procedure defined by Nistor et al. [2]. The dependent variable was *community response*, defined on a scale with five values: 1 = no reaction, 2 = deleted comment, 3 = approved comment (in moderated blogs), 4 = single answers, 5 = discussion on visitor's inquiry. This was regarded as an interval scale.

The blogger communities were randomly selected from the Internet. One of the researchers posted the inquiries according to the experimental design. Subsequently, the community response was followed and recorded for the two weeks after the inquiries had been posted. The comments posted during the entire lifetime of each blog were locally saved, and analyzed with the ReaderBench tool [2, 9]. Finally, all collected data were statistically processed employing IBM SPSS Statistics version 22 for Mac.

4 Findings

The measured *collaborative dialog quality* indicators varied within very large limits and were strongly correlated with each other. Hence, they were normalized (z values were calculated) and a main component analysis with oblimin rotation was performed. Two components were extracted (by means of the Anderson-Rubin method) and used in the further analysis to describe the collaborative community dialog quality. These components explained 79 % of the total variance. The first factor (in the following referred to as 'overall discourse quality') corresponds to the dialog quality in the entire community's discourse, including the voice inter-animation. The second factor (referred to as 'comments quality') corresponds to the dialog quality in single comments. No significant differences were found between blog topics in terms of overall discourse quality. A marginally significant difference was found in the comments quality ($F = 3.98, p = .050$), such that the comments quality was higher in politics & economy ($M = .26, SD = 1.36$) than in cooking blogs ($M = -.22, SD = .47$).

To determine the *socio-cognitive structure* of the examined OKCs, a similar dimension reduction procedure was performed for each community discourse. Subsequently, the participants of each community were hierarchically clustered according to the Ward method, extracting three clusters that corresponded to the community layers: central participants ($M = 3, SD = 11$), active participants ($M = 159, SD = 292$) and peripheral participants ($M = 1322, SD = 319$).

To answer RQ1, the *effects of the socio-cognitive structure on the established dialog quality* in the OKCs were tested by means of multiple regression analysis. The overall discourse quality was significantly impacted by the number of active members ($\beta = .51, p < .000$), peripheral members ($\beta = .47, p < .000$) and central members ($\beta = .25, p < .01$), which cleared $\text{corr. } R^2 = .55$ of the total variance in overall discourse quality.

Community response covered the entire scale range from 1 to 5, with $M = 2.76$ and $SD = 1.34$. To answer RQ2, a variance analysis (UNIANOVA) corresponding to the 2×2 factorial design, was performed. This revealed a strong, significant main effect of the visitor inquiry format: $F(1, 65) = 39.83, p < .000$, $\text{partial } \eta^2 = .41$, explaining $\text{corr. } R^2 = .45$ of the total variance in community response. The blog topic effect, the interaction and covariate effects were not significant.

Additionally, these effects were examined separately, by multiple regression analysis in the two blog topic groups. For cooking blogs only, the community response was

strongly impacted by the visitor inquiry format ($\beta = .72, p < .000$). This model explained $\text{corr. } R^2 = .62$ of the total variance. For politics & economics blogs, the community response was only influenced by dialog quality, i.e., by the percentage of social knowledge building per utterance scores ($\beta = 1.00, p < .05$), which explained $\text{corr. } R^2 = 0.43$ of the total variance. In this case, neither the visitor inquiry format, nor the socio-cognitive community structure had significant effects on the OKC response.

5 Discussion and Conclusions

This study examined the effects of visitor inquiry format, topic, established collaborative dialog quality and socio-cognitive structure on the response of a blogger OKC. As expected, the visitor inquiry format had a significant impact on community response, meaning that OKCs were more likely to respond to on-topic than to off-topic inquiries. This relationship was connected with the established blog topics. In the examined cooking blogs, the community response was strongly influenced by the inquiry format only. In contrast, in the politics & economics blogs, the community response depended only on the collaborative dialog quality, i.e., on social knowledge building. This difference may be caused by a difference in the specific knowledge construction process and employed communication strategies, which in turn may be given by the difference in topic. While discussing politics and economy, OKC members are likely to elaborate more, and create more knowledge [3] than while exchanging cooking recipes. Consequently, the success of a visitor inquiry may be a matter of netiquette in cooking blogs (i.e., of on-topic comments), and a matter of knowledge and argumentation (i.e., inter-animation of voices [5]) in politics & economy blogs.

Remarkably, the community structure impacted the overall dialog quality, but it had no direct effect on the community response. This seems to contradict the findings reported by Nistor and colleagues [2], who provided empirical evidence for a significant effect of socio-cognitive structure on OKC response.

The results of this study suggest a shift in the initial research model. Initially, it was assumed that the visitor inquiry format, the blog topic, the established collaborative dialog quality, and the socio-cognitive OKC structure of the examined communities would have direct effects on the community response. The presented results suggest that only two factors may have direct effects on community response, namely the visitor inquiry format, and the established dialog quality. The latter may be further impacted by the socio-cognitive OKC structure, i.e., collaborative dialogue quality may mediate the effect of socio-cognitive structure on community response. Such a model would reconcile the findings reported above with those reported by Nistor and colleagues [2]. This would be worth verifying in follow-up research.

For the practice of technology-enhanced learning, this study suggests that social learning analytics can support informal learning in OKCs by automatically assessing the quality of the collaborative dialog in various OKCs, and subsequently recommending those with higher dialog quality as more likely to be responsive and integrative communities [2, 13]. At the same time, OKC visitors should have a minimum of insight in community's knowledge and "ways of doing things" [4], observing the interaction rules and strategies that are specific to the chosen OKC.

Acknowledgements. This work has been partially funded by the 644187 RAGE H2020-ICT-2014 project, as well as by the Sectorial Operational Program Human Resources Development 2007-2013 of the Romanian Ministry of European Funds according to the Financial Agreements POSDRU/159/1.5/S/134397 and 134398.

References

1. Baker, R., Siemens, G.: Educational data mining and learning analytics. In: Sawyer, K. (ed.) *Cambridge handbook of the learning sciences*. Cambridge University Press, Cambridge (in press) (2014)
2. Nistor, N., Dascălu, M., Tarnai, C., Bresser, N., Trăușan-Matu, S.: Online knowledge communities as student-centered open learning environments: how likely will they be to integrate learners as new members? In: 11th International Conference on Computer-Supported Collaborative Learning (CSCL 2015). ISLS, Gothenburg, Sweden (in press) (2015)
3. Bereiter, C., Scardamalia, M.: Knowledge building and knowledge creation: one concept, two hills to climb. In: Tan, S.C., So, H.J., Yeo, J. (eds.) *Knowledge Creation in Education*, pp. 35–52. Springer, New York (2014)
4. Lave, J., Wenger, E.: *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, Cambridge (1991)
5. Bakhtin, M.M.: *The Dialogic Imagination: Four Essays*. The University of Texas Press, Austin and London (1981)
6. Trăușan-Matu, S., Stahl, G., Sarmiento, J.: Polyphonic support for collaborative learning. In: Dimitriadis, Yannis A., Ziguers, I., Gómez-Sánchez, Eduardo (eds.) *CRIWG 2006. LNCS*, vol. 4154, pp. 132–139. Springer, Heidelberg (2006)
7. Trăușan-Matu, S.: The polyphonic model of hybrid and collaborative learning. In: Wang, F.L., Fong, J., Kwan, R.C. (eds.) *Handbook of Research on Hybrid Learning Models: Advanced Tools, Technologies, and Applications*, pp. 466–486. Information Science Publishing, Hershey, NY (2010)
8. Dascălu, M., Trăușan-Matu, S., Dessus, P., McNamara, D.S.: Discourse cohesion: a signature of collaboration. In: 5th International Learning Analytics & Knowledge Conference (LAK 2015). ACM, Poughkeepsie, NY (in press) (2015)
9. Dascălu, M.: Analyzing discourse and text complexity for learning and collaborating. *Studies in Computational Intelligence*, vol. 534. Springer, Switzerland (2014)
10. Dascălu, M., Trăușan-Matu, S., Dessus, P., McNamara, D.S.: Dialogism: a framework for CSCL and a signature of collaboration. In: 11th International Conference on Computer-Supported Collaborative Learning (CSCL 2015). ISLS, Gothenburg, Sweden (in press) (2015)
11. Bakhtin, M.M.: *Problems of Dostoevsky's Poetics*. University of Minnesota Press, Minneapolis (1984)
12. Schmidt, J.: Blogging practices: An analytical framework. *J. Comput. Mediated Commun.* **12**(4), 1409–1427 (2007)
13. Nistor, N., Trăușan-Matu, S., Dascălu, M., Duttweiler, H., Chiru, C., Baltes, B., Smeaton, G.: Finding student-centered open learning environments on the internet: Automated dialogue assessment in academic virtual communities of practice. *Comput. Hum. Behav.* **47**(1), 119–127 (2015)