Newcomer Integration
in Online Knowledge Building Communities:
Automated Dialogue Analysis
in Integrative vs. Non-Integrative Blogger Communities

Nicolae Nistor[1,2,3], Costin-Gabriel Chiru[4], Nicolas Bresser[1]

[1] Faculty of Human Sciences, Universität der Bundeswehr München
Werner-Heisenberg-Weg 39, D-85577 Neubiberg, Germany, nic.nistor@unibw.de
[2] Faculty of Psychology and Educational Sciences, Ludwig-Maximilians-Universtität
Leopoldstr. 13, D-80802 München, Germany, nic.nistor@uni-muenchen.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, RO-60042 București, Romania, costin.chiru@cs.pub.ro

Abstract. Online knowledge building communities (OKBC) reunite participants engaged in collaborative discourse. OKBCs can be made „smart" by adding tools that predict how likely an OKBC is to integrate newcomers in existing dialogues and socio-cognitive structures. Starting from Bakhtin's dialogical approach and polyphony theory, and building on the concept of inter-animation of voices, this study explores the relationship between newcomer integration and dialogue quality in OKBCs. The automated analysis tool "Important Moments" was employed to compare two dialogues, from an integrative and from a non-integrative blog-based OKBC. In the former, the concepts, lexical chains and inter-animation moments occurred more frequently than in the latter. Also, newcomer comments were linked to less lexical chains in the integrative community than in the non-integrative OKBC. These findings suggest close relationships between dialogue quality and newcomer integration, which can be used for designing smart OKBCs.

Keywords: Online communities, Newcomer integration, Community discourse, Inter-animation of voices

1   Introduction

Online knowledge building communities (OKBC) [1], [2] concentrate intellectual capital by reuniting participants who engage in social behaviors including situated, problem-based collaborative discourse, and the construction of applicable knowledge and skills [3, 4]. Technology can be employed to make OKBCs "smart", which results in a better "flow" of the socio-cognitive activities [4]. In such applications, technology is expected to empower people in their social practice by monitoring the socio-cognitive activity, providing appropriate feedback, predicting further evolution [5], and thus enabling the co-construction and the convergence of knowledge [4], [6].

People – no matter if community members or outsiders – may seek for answers to their questions and solutions to their problems in online discussions. Thus, they attempt to participate in thematically related community discourse, and to enter the community carrying out the discourse. This in turn poses the question of integrativity: How likely will an OKBC be to integrate new members in the ongoing discourse, and in the existing socio-cognitive structures?

A deeper understanding of newcomer integration [7] requires first an examination of the community phenomenon. The particular "ways of doing things" [8] includes a communication style and discourse that are characteristic for a given community, and that are tightly related to participation patterns and member identity in that community. Hence, there may be a close relationship between community discourse and the integration of newcomers, a phenomenon insufficiently studied so far [7]. In other words, the community discourse may indicate how likely a community may be to integrate newcomers. Consequently, the corresponding discourse analysis may predict integrativity, support productive collaboration, and thus implicitly foster co-construction and convergence of knowledge.

This paper presents a discourse analytics approach and an automated tool called „Important Moments", both aimed at assessing the interanimation of voices in text-based discussions. In a subsequent case study, conclusions about OKBE are drawn from a comparison between analysis results in an integrative and a non-integrative OKBC.

2   Theoretical Background

2.1   Discourse in Online Knowledge Building Communities

Knowledge building communities (KBCs) are groups of mutually engaged people sharing interests and goals, while constructing shared knowledge and "ways of doing things" [1], [8]. Participation and knowledge construction are essential parts of a community discourse that can be sustained via various communication channels, including online communication in the case of OKBCs.

KBCs are characterized by socio-cognitive structures integrating participants with different degrees of participation, expertise levels and status [8]. Central participants perform more difficult tasks and assume more responsibility than peripheral members. This requires a higher level of expertise for central participation, which includes domain knowledge and experience. Participation in community discourse is an expression and application of individual expertise, which in turn may lead to recognition of an expert status [9]. In other words, expert status is negotiated in the community discourse [8], as well as community membership.

Both community membership and status within the community are prone to changing in time. Ideally, newcomers start as visitors and legitimate peripheral members and evolve towards central participation. On the other hand, members can lower their participation or even leave the community [8]. Therefore, a community can survive in the long term only if it continuously integrates new members. Newcomers' access to communities, and to community knowledge, may be open or

restricted. Community practices include new member recruiting, assessment and acceptance strategies and procedures [2], [7], as well as participation support structures. From such structures, the recruitment strategies and the accessibility of community knowledge are predictive of the newcomers' level of participation [7].

While Jean Lave and Etienne Wenger, the authors of the community of practice approach have asserted from the very beginning that membership, practice and discourse are tightly related to each other [8], little is known about the mechanisms of newcomer integration, and about the ways how these are reflected in the community discourse. Recent research results [7] suggest, however, a close relationship between community discourse and the integration of newcomers. In other words, the community discourse may indicate how likely a community may be to integrate newcomers, and how the integration may take place.

2.2 Automated Discourse Analysis

A possible approach to community discourse is defined by Bakhtin's dialogism [10]. The central piece of this approach is the concept of voice, a standpoint taken by one or more participants, discussed in the dialogue, and influencing the subsequent evolution of the dialogue [11]. Thus defined, a participant can have several voices, and a voice can be uttered by many participants. Voices are built of repetitions of words expressing the standpoint in question [12]. In a strict sense, words can be repeated in the same form, as a repetition chain. In a larger sense, repetitions can also include synonyms, paronyms, and further related words, building thus lexical chains. Both repetition chains and lexical chains [13] can be identified using ontologies or lexical databases, and are useful for the identification of voices in dialogues [9].

According to Bakhtin [10], any text is a dialogue with a meaning given by the relationships between voices. This means that the participating voices influence each other similarly to the themes and voices in polyphonic music. Bakhtin describes this phenomenon as inter-animation of voices. Chiru and Trăușan-Matu [14] describe several types of inter-animation points between voices:

- Pivotal moments: Two voices meet in the same utterance, and one voice changes into the other.
- Convergence moments: Two or more voices meet in the same utterance, and all of them disappear from the dialogue.
- Singular moments: Two or more voices meet in the same utterance, one of them continues to be present in the dialogue, and all others disappear.
- Divergence moments: Two or more voices meet in the same utterance, after which each of them continues its flow in the dialogue.
- Meeting points are chains of inter-animation moments (mostly divergence) in which voices meet several times in the discourse.

In a study of chat-based dialogue, Chiru and Trăușan-Matu [14] found a positive correlation between the number of inter-animation moments and dialogue quality as evaluated by experts. The evaluation criteria were both content and participant involvement. The latter implicitly includes the integration of community members, meaning that an integrative community is likely to sustain a dialogue in which inter-

animation moments occur more frequently. However, available empirical evidence is insufficient to sustain this assumption.

2   Case Analysis

This study explores the relationship between newcomer integration in OKBCs and dialogue quality regarded as inter-animation of voices. Since this line of research is at the very beginning, the study considers a small number of cases and aims at generating hypotheses for further research.

2.1   Setting and Choice of Cases

The study was carried out in OKBCs emerging from blogs that were freely available in the Internet on various platforms, such as blogspot.com and wordpress.com. The integrativity information about the studied OKBCs was gathered in a prior experiment in the following manner. While conducting an internet-based questionnaire survey for another study, the researchers had visited these OKBCs and invited the members to respond to the survey. The response rate was quite diverse; some OKBCs proved *integrative* by responding the survey, others proved *non-integrative* by ignoring the interventions of the researchers. This experience was used in this case study.

From these OKBC categories, the researchers randomly chose for the analysis two blog-based dialogues, one from an integrative and the other from a non-integrative OKBC. The initial posting from the OKBC described blog owner's encounters with her students in the first week of the school term. The initial posting from the non-integrative OKBC described blog owner's encounters with family members at a meeting abroad. Extracts of the two dialogs consisting of the first five comments of each are provided in the Appendix along with the complete URLs.

2.3 Analysis Tool

In the following, the tool "Important Moments" [14] was employed. Initially developed for synchronous chats, this tool analyses text files containing dialogues, follows the existing voices, and detects the points of interaction between the voices that are present in those dialogues. The output includes numeric descriptive data such as the number of participants, utterances, words, concepts, voices, repetition and lexical chains, and inter-animation moments. Moreover, it provides a visual representation of the dialogue, voices and inter-animation points. The tool identifies the inter-animation moments based on the assumption that in order to influence each other, the voices need to be present in the same area of text. (There is a higher probability that a voice influence other voices that are found in its context than those which are not found close to the given voice). For most dialogues, the voices are considered to interact if they are found in the same utterance. However, this definition of context (for a given voice) is flexible, depending also on the dimension of the analyzed dialogue. If the dialogue is extremely large, then multiple utterances are

considered as context instead of a single one. After recognizing the inter-animation moments, the tool follows the voices and identifies the type of interaction that occurred.

2.3 Findings

The presentation of findings begins, for better comprehension, with examples of inter-animation moments that were identified in the two dialogues. In Fig. 1, a pivotal moment is depicted representing the passing from one idea to another. The voice of "start" dies out and a new one ("great") appears in the dialogue. In order to find out what the conversation was about in the moment of the interaction, a click on the triangle will reveal the text corresponding to that moment, along with the context of its appearance (Fig. 2).

Fig. 1. Example of a pivotal moment in the dialogue of the integrative OKBC
Full circle = concept being introduced; triangle = pivotal moment; each color represents a different lexical chain

Fig. 2. Example of the text that generated the pivotal moment from Fig. 1.

Fig. 3 shows a convergence moment between the voices of "time" and "family". After these two meet several times during the dialogue (creating a chain of meeting points), they have a final meeting in the end and after that both voices die out. This moment can be interpreted as reaching a conclusion after which nothing is worth saying, or as a reconciliation of a dispute between two or more voices. The text corresponding to this point is presented Fig. 4.
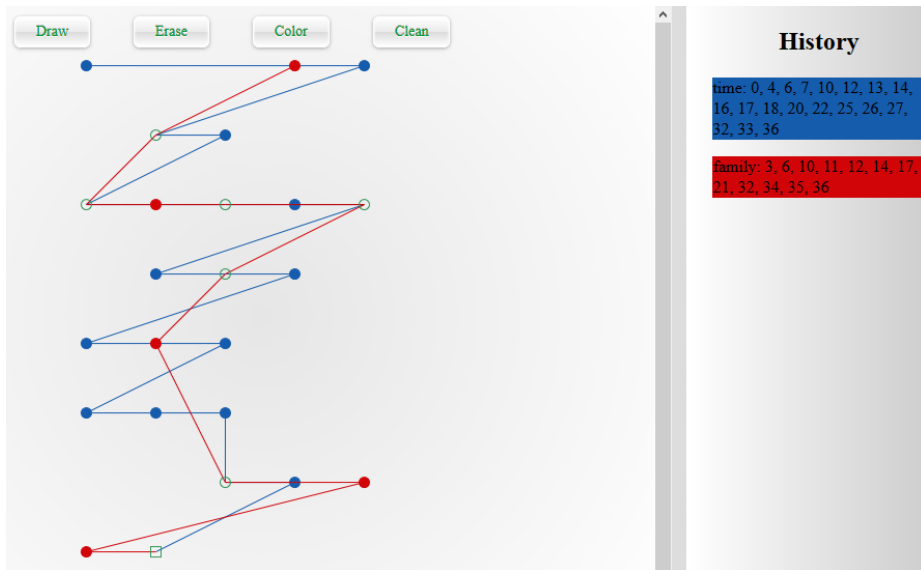
Fig. 3. Example of a convergence moment in the dialogue of the non-integrative OKBC
Full circle = concept being introduced; empty circle = meeting point; square = convergence
moment; each color represents a different lexical chain

Fig. 4. Example of the text that generated the convergence moment from Fig. 3

The final type of moment considered in this research is the singular moment. An example is provided in Fig. 5, where the voices of "tips" and "post" meet in the 35th utterance of the dialogue (in Fig. 5 the utterance number is 34, but the counting starts from 0, not from 1). After that, only the voice of "tips" continues to be present, the other one being diluted. This type of moments can be interpreted as a fight between two or more voices, where only the strongest one survives. The corresponding text is presented in Fig. 6.
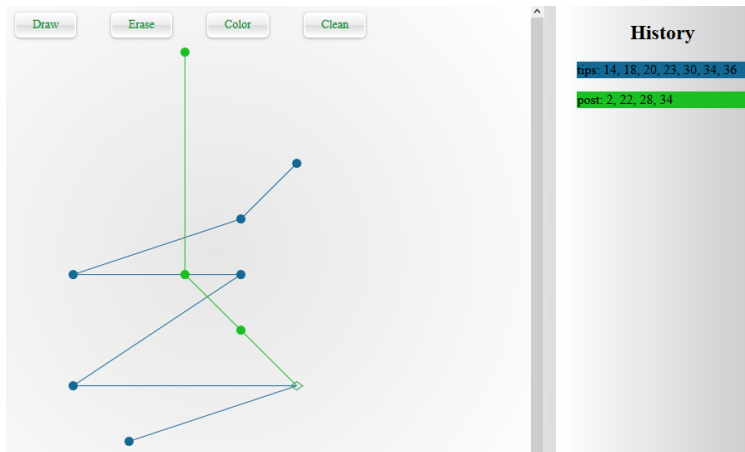
Fig. 5. Example of a singular moment in the dialogue of the non-integrative OKBC
Full circle = concept being introduced; diamond = singular moment; each color represents a
different lexical chain

Fig. 6. Example of the text that generated the singular moment from Fig. 5

In sum, the analyzed dialogues display the numbers of inter-animation points presented in Table 1. In the integrative OKBC, less bloggers participated; therefore the discussion consisted of less but longer comments, as compared with the discussion from the non-integrative OKBC. The discussion addressed a smaller number of concepts in total (142 vs. 344), however more concepts per 1000 words (347 vs. 240); and, similarly, a smaller number of repetition chains in total (6 vs. 11), but twice as many repetition chains per 1000 words (14.7 vs. 7.6). The inter-animation moments occur (per 1000 words) more frequently in the integrative than in the non-integrative OKBC: 696.8 vs. 0.7 pivotal moments; 1408.3 vs. 483.9 convergence moments, and 1442.5 vs. 144.8 singular moments. (Inter-animation moments are calculated based on combinations of words, therefore a number of moments greater than the total number of words is possible.) Since in both OKBCs the numbers of divergence moments were negligible, and the numbers of meeting points very high, both were disregarded. Both discussions displayed similar distribution patterns of inter-animation moments, comprising meeting points in the entire dialogue, pivotal moments in the second half of the dialogue, singular and convergence moments towards the end of the dialogue (Fig. 7).

Table 1. Comparison of blog discussions from the integrative vs. non-integrative OKBC

	Integrative OKBC	Non-integrative OKBC
Number of participants		
• total number	5	20
• per 1000 words	12.2	14.0
Number of comments	8	37
Average comment length (words/comment)	52	40
Total number of words	409	1430
Number of concepts		
• total number	142	344
• per 1000 words	347	240
Repetition chains		
• total number	6	11
• per 1000 words	14.7	7.6
Pivotal moments		
• total number	285	1
• per 1000 words	696.8	0.7
Convergence moments		
• total number	576	692
• per 1000 words	1408.3	483.9
Singular moments		
• total number	590	207
• per 1000 words	1442.5	144.8
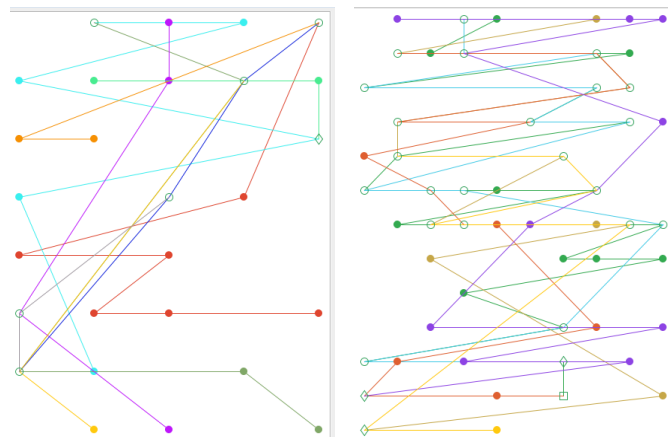
Fig. 7. Repetition chains and inter-animation moments in the dialogue of the integrative OKBC (left) vs. non-integrative OKBC (right)
Full circle = concept being introduced; empty circle = meeting point; diamond = singular moment; square = convergence moment; each color represents a different lexical chain

Additionally, a comment made by a visitor of the non-integrative OKBC was compared with a comment made by a visitor of the integrative OKBC with respect to the integration of these comments in the dialogue. With this purpose, the links between each comment and the repetition and lexical chains of the entire discussion

were checked (Table 2). While both comments had the same length (50 words) and were linked to the same number of repetition chains (11), the intervention in the integrative OKBC was more focused, and linked to less lexical chains (3 vs. 4).

Table 2. Comparison of blog visitor comments from the integrative vs. non-integrative OKBC (bold words were checked for links, words in italics were linked in repetition chains, colored words were linked in lexical chains; each color represents a different lexical chain)

		Words	Repetition chains	Lexical chains
Integrative OKBC	Miss.Kay: I saw your *first day activity* from *last year* and **definitely** "**stole**" it. I *loved* how it **gave** the **students** the **chance** to **tell** me what they needed. I **already** referred to it this **year**. I am a **sixth** *grade* **co-taught** **math** teacher and I look **forward** to your posts.	50	11	3
Non-integrative OKBC	Nicolas: *Great tips*. I'm from *germany* and one part of my *family* from *england*. So *planning* **time** together is very important when they come for a *visit*. *Planning* for two (aunt and uncle) was allready *difficult*, so *planning* for 4... respect. *hope* there will be **more** from you. Greets from *Germany*	50	11	4

3 Discussion

Comparing the two blog discussions, some similarities appear that may be general characteristics of casual chat dialogues: opening and sustaining the dialogue by introducing concepts, repeating themes in lexical chains, and producing meeting points; closing the dialogue by pivotal and convergence moments. However, this distribution may be different in more productivity-oriented types of dialogue, such as in brainstorming for instance.

There were also considerable differences between the analyzed blog discussions. The dialogue from the integrative OKBC had more repetition and lexical chains, and more inter-animation moments. This suggests that the discourse of integrative OKBCs may include more elements of inter-animation than in non-integrative OKBCs. These findings are consistent with the results reported by Chiru and Trăușan-Matu [14] that high quality dialogue is associated with a higher number of important moments.

4 Conclusions

Of course, the validity of the presented results is limited by the small number of analyzed cases. To overcome this limitation, the assumption of the correlation between integrativity and number of interanimation points in the discourse should be tested with a larger number of analyses and more diverse dialogues.

If this assumption proves true in a context larger than the case study presented above, a practical conclusion is straightforward. When searching the Internet for OKBCs, it is a simple operation to use the "Important Moments" tool to assess the frequency of inter-animation points and predict OKBC integrativity. Existing OKBCs can be made "smart" by adding a tool such as "Important Moments".

For educational research, this study contributes to uncovering the relationship between newcomer integration [7] and discourse in online communities [1, 2, 3, 8, 11, 12, 14]. Also, it makes headway in establishing Learning Analytics methodologies [5] that support educational decisions related to open collaborative learning environments [2].

References

1. Bereiter, C., Scardamalia, M.: 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, Springer, New York, (2014)
2. Nistor, N.: Knowledge communities in the classroom of the future, in Classroom of the future: Orchestrating collaborative spaces, K. Mäkitalo-Siegl, F. Kaplan, J. Zottmann. & F. Fischer, Eds., pp. 163-180, Sense, Rotterdam, (2010)
3. Granovetter, M. S: The Strength of Weak Ties. Am. J. Sociol. 78, pp. 1360–1380 (1973).
4. Giovannella, C., Dascalu, M., Scaccia, F.: Smart City Analytics: state of the art and future perspectives. Interact. Des. Archit. J. 20, 72–87 (2014).
5. Baker, R., Siemens, G.: Educational data mining and learning analytics, in Cambridge handbook of the learning sciences (2nd edition), R. K. Sawyer, Ed., Cambridge, UK: Cambridge University Press, (in press) http://www.columbia.edu/~rsb2162/ BakerSiemensHandbook2013.pdf
6. Murillo Montes de Oca, A., Nistor, N., Dascălu, M., Trăușan-Matu, Ș.: Designing smart knowledge building communities, in this special issue.
7. Eberle, J., Stegmann, K., Fischer, F.: Legitimate peripheral participation in communities of practice: Participation support structures for newcomers in faculty student councils. Learning and Instruction (in press)
8. Lave, J., Wenger, E.: Situated learning. Legitimate peripheral participation, Cambridge: University Press, (1991)
9. Nistor, N., Fischer, F.: Communities of practice in academia: Testing a quantitative model. Learning, Culture and Social Interaction, 1(2), 114–126 (2012).
10. Bakhtin, M. M.: The dialogic imagination. Four essays, Austin: University of Texas Press, (1981)

11. Trăuşan-Matu, Ş.: The polyphonic model of hybrid and collaborative learning, in Handbook of research on hybrid learning models: Advanced tools, technologies, and applications, in F. L. Wang, J. Fong, & R. C. Kwan, Eds., pp. 466--486, Hershey, PA: Information Science Publishing, (2010)
12. Tannen, D.: Talking voices: Repetition, dialogue, and imagery in conversational discourse, Cambridge University Press, (1989)
13. Morris, J., Hirst, G.: Lexical cohesion, the thesaurus, and the structure of text, Computational Linguistics, 17(1), pp. 211-232 (1991)
14. Chiru, C. G., Trăuşan-Matu, Ş.: A tool for discourse analysis and visualisation, International Journal of Virtual Communities and Social Networking, 5(2), pp. 55-71 (2013)

Appendix

Dialogue Extract from the Integrative OKBC

```
http://7thgrademathteacherextraordinaire.blogspot.de/2013/09/first-day-
of-school.html
```

Miss.Kay: I saw your first day activity from last year and definitely "stole" it. I loved how it gave the students the chance to tell me what they needed. I already referred to it this year. I am a sixth grade co-taught math teacher and I look forward to your posts.

Sherrie Nackel: Hi Miss Kay, I got the idea from someone last year. Glad you could use it. I do love the feedback I get so quickly and easily. I have many more posts coming about my first week once things start to settle down here.

Melanie Bates: Sherrie, What did you use for your "chart paper"? It doesn't look like chart paper. Almost looks like large whiteboards. Melanie

Sherrie Nackel: Hi Melanie, That was just tagboard I wrote the prompt with Crayola poster markers and then laminated. I will use them again next year, that's why I laminated them. Hope your year is off to a great start.

Lindaisy: Hi Sherrie! I've read a lot of your different blog posts and I subscribe too I wanted to ask you something about your homework/classwork grading methods. You state that you don't collect daily assignments, so what do you actually physically collect and then grade? Do you keep their tests and quizzes in a file in your room? Also, when you put grades in, are you putting 2pts as their homework grade each time. For instance, 2/2. I'm just curious because I'm implementing something similar but when it comes time to physically put their grades in the computer, I felt funny putting 2/2 as a grade! Please respond. Thanks. LOVE your blog!

Dialogue Extract from the Non-Integrative OKBC

```
http://lostintravelsblog.com/2013/11/survive-a-visit-from-your-in-laws/
```

Esther: i think taking trips with in-laws is even harder when we're all out of our comfort zones. i'm pretty lucky to have great in-laws too.. but yeah.. scheduling time alone.. i think that's what i need to remember next time!

Chelsea: ha! i couldn't agree more! thankfully for us, even when we all visited a different country together, things ran rather smoothly

Katie: I'm just saying, this was SUCH a great post! You are so good at thinking up these great ideas:) Glad you guys had a blast!

Chelsea: haha glad that it could help! maybe you could use some of these when you go visit family for christmas this year!

Rachel: It can be so difficult to spend that much time with other people. Alone time is a great suggestion to retain your sanity! And I really like the idea of doing something outside of their comfort zone