Reviewer’s report

Title: How the Study of Online Collaborative Learning Can Guide Teachers and Predict Students' Performance in A Medical Course

Version: 1 Date: 13 Jun 2017

Reviewer: Ingo Kollar

Reviewer's report:

The manuscript has substantially improved from the previous to the current version. The authors now clearly demonstrate the potentials that SNA offers when it comes to the analysis of learning processes on different levels (course level, thread level, individual level) and to the explanation of external performance criteria. They also did overall a good job in attributing SNA as a method itself explanatory value and now instead convincingly argue that SNA is one more tool in the researcher's hands to identify processes of collaborative learning that eventually lead to individual performance.

While my overall impression of the manuscript is positive, some minor issues remain that should be addressed in one further revision of the paper:

1. Abstract: "and the relationship of SNA with students' performance remains largely unknown" --> how can a research method have a relation with students' performance? From my point of view, this sentence should read somehow like this: "and the relationship of parameters that can be obtained via SNA with students' performance remains largely unknown".

2. Abstract (results): "the interaction patterns and the significant mediators" --> mediators between what variables? Please explain.

3. P. 5, line 37: "...see humans are..." needs to be changed into "...see humans as..."

4. P. 6: You are critizing dashboards that are usually used in LMSs for their restricted possibility in data analysis. I wondered whether you could give an example of such an LMS and the analytic dashboard it uses. Perhaps you could even present a screenshot that makes its restrictions visible?

5. Maybe my most important remark: I still think that the different SNA parameters you presented could be better introduced. I think the reader would benefit much from an explanation of how each of the parameters are computed when these explanations are already presented in the theoretical section, i.e. on p. 7 and 8. E.g. at the bottom half of p.8, you say what "centrality" means, but it remains unclear how it is computed. The same applies for the
introduction of all other parameters in that section ("closeness centrality", "weighted degree centrality" etc.).

6. Also on p.7, you say that SNA offers you to deliniate the "importance" in mediating interactions. Here I wondered what you exactly mean by "importance". At first sight, I would understand "importance" (e.g. of a contribution) as something that requires content-analytical techniques to assess. You however seem to mean something else (I assume the importance of a person as indicated by in- and out-degree centrality). Please explain in a clearer way what you mean here.

7. I very much appreciate that the authors ask clear research questions. Yet, the results section could do a better job in describing what analysis refers to what research question. This could be done by using appropriate headings such as "The social structure and performance (research question 2)" and by possibly repeating each research question in the results section before the respective analyses are presented.

8. Research question 3 seems to be broader than the analyses you ran. It reads "How can student's position, interactions, and relations in a network be used to predict his or her final performance?". In the analysis, you however restrict that question to the identification of underachievers (which I - by the way - do not understand how you identified them -should be described as well). I would thus recommend to be more specific in the formulation of the research question, i.e. to restrict it to the identification of underachievers.

9. P. 13: I do not quite understand why "role in moderating and relay of information" is bolded and why it is presented before the parameter (as this is always done the other way round in case of the other parameters that are presented in bullet points”.

10. P. 13: I am still not completely sure if I understood correctly what you mean by betweenness centrality (does this mean the frequency by which a person has received information from one participant and passed it on to other participants?).

11. P. 13: Could you add information on how information centrality and closeness centrality are actually computed (right now, you rather abstractly talk about those two concepts). The same goes for "clustering coefficient".

12. P. 14 (bottom): please add information on what research questions are addressed by (a) network visualization and (b) network analysis.

13. Figure 3b: I am still not sure if I understand the "out-degree centrality" figure. What exactly is the difference to figure 3a? From my understanding, 3a shows that the teacher received a lot of messages from many students; and in 3b, I take that students give a lot of information to the teacher. So what is the difference between the two visualizations, also given the fact that they look very similar to each other?
14. P. 20: Again, please add what research question you tackle under the heading "network properties and correlation to performance".

15. P. 20: perhaps I missed it - but if not: please explain what you mean by "graph density", "average degree" and "average clustering coefficient".

16. P. 22: Again, please add what research question you tackle under the heading "Can social network analysis predict performance" (and as recommended above: please check whether this question could be formulated in a more specific way --> underachievers).

17. As I am not familiar with ALM, I was wondering whether its results should include more information than just the amount of explained variance. What about beta-weights and p-values in tables 3 and 4? At least for "regular" regression, I would expect this information to be presented.

18. Table 5: Maybe I missed it, but did you explain how you identified students as "at-risk" or "safe"? Based on the grades in the final exams?

19. P. 27, second paragraph: Here you start the discussion saying "On the course level..." - I was subsequently missing sentences that would start with "On the thread level" and "on the individual level", i.e. you could perhaps go through these three levels in a more systematic way.

20. Although the English is very good, especially the newly added or revised parts need some additional language editing (e.g., I see articles such as "a" or "the" missing here and there).

**Are the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

No

**Does the work include the necessary controls?**
If not, please specify which controls are required in your comments to the authors.

Yes

**Are the conclusions drawn adequately supported by the data shown?**
If not, please explain in your comments to the authors.

Yes

**Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?**
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.
I am able to assess the statistics

**Quality of written English**
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license ([http://creativecommons.org/licenses/by/4.0/](http://creativecommons.org/licenses/by/4.0/)). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal