Author’s response to reviews

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

Authors:

Mohammed Saqr (saqr@dsv.su.se)

Uno Fors (uno@dsv.su.se)

Matti Tedre (matti@dsv.su.se)

Version: 1 Date: 27 May 2017

Author’s response to reviews:

Dear Editor,

Thank you for the opportunity to revise and improve our manuscript. We have tried to fully address all comments from the reviewers and have done the necessary changes that we believe have significantly improved the manuscript.

We believe that our study presents a novel analytical technique to better understand online collaborative learning in a medical education setting. Our study is probably the first to address the usage of SNA indicators in early prediction of students’ achievement, the use of information centrality as a predictor, the predictability of clinical performance, the role of dynamic networks in education and the study of clinical case discussions, and how that can possibly inform intervention.

Most reviewer comments have centered on two major points

• Reviewers recommended to expand the section on the predictive potential of Social network analysis.

  o So, we have re-written the last three pages of the introduction to discuss the literature on the subject of using SNA in learning analytics. We have also improved the methods section by adding more details to the SNA measures used in the study, description of the predictive algorithm and the statistical tests.
We have expanded the analytics results by including data for early participation (correlation and prediction of midterm and final performance and used logistic regression to investigate the possibility of early prediction of under acheivers.). Reviewer 2 recommendation.

We have also expanded the final performance results by including results of clinical performance, knowledge and comprehension (MCQ) and final grades. Reviewer 3 recommendation.

These findings were further discussed in the discussion section.

- It seems that some phrases in the original manuscript suggested that our position is that SNA derived parameters are more important than individual attributes. We have removed all these passages and made it clear that they are both inter-related and our work is to demonstrate the potential of this technique in offering some insights into the online collaborative learning process.

Three other points worth mentioning:

- We have removed two network parameters (PageRank and Harmonic closeness centrality) since other network parameters are very similar to them (Eigen and closeness), to decrease complexity recommended by reviewer 2. The analysis was repeated, the model accuracy slightly changed since these two parameters were excluded.

- The correlation was re-done using Kendall’s Tau-b per reviewer 2 recommendation, so some previous correlation coefficient results were slightly different using the new test.

Below, our detailed answers to all of the comments by the reviewers can be found.

We have highlighted the changes in blue (added text) or brown (mostly re-written and improved)

We hope that with these changes and updates, that the manuscript is acceptable for your journal.

Sincerely yours,

Mohammed Saqr
Reviewer reports:

John Sanders (Reviewer 1): The topic is of potential interest to a wider audience where there is use of online collaborative learning in medical education but I have several recommendations:

1. It is important in the title and abstract to clearly state the type of social networks that are being analyzed - these are online collaborative learning in a formal module.

Both title and abstract were modified to reflect this, the type was made clear in the background, aim, methods, the results, and conclusions

2. There is a discussion of the use of social network analysis in medical education but it is essential to discuss the type of social network - online etc.

The type of social network was further clarified and explained in the context section, (P10). Sections of introduction and discussion were improved and this has been discussed further in a clearer way.

3. The use of social network analysis is within the wider use of learning analytics but this is not discussed and it is essential to relate this study to the wider literature on learning analytics and also the previous use of social network analysis for online collaborative learning.

Page 8,9 were re-written, learning analytics was discussed in detail, with focus on SNA and how it was used in previous learning analytics research. It has also been further discussed in the discussion and relevant sections throughout the manuscript.

We have expanded the section on the predictive potential of Social network analysis and expanded the results by including data for early participation (correlation and prediction of midterm and final performance).

We have also expanded the final performance results by including results of clinical performance, knowledge and comprehension (MCQ) and final grades.

These findings were further discussed in the discussion section.

4. The social network data collection and analysis is clearly presented but I recommend some further discussion of the justification or validity of the approach that was used.
The centrality measures were re-structured, further explained (P13-14), examples and rationale were discussed from the previous literature under the heading network analysis (P8-9). For clarification, we grouped importance of contribution into three main roles (Quantity of activity, Role in information relay and strength of connectedness)

5. The conclusions are appropriate but it is important that all of the figures have clear captions and labels to describe what they are presenting to the reader.

Figure caption were revised and clarified.

6. Learning analytics and social network analysis present large amounts of data that can be interpreted but some discussion of the practical potential uses would help the reader to consider its application eg offering support etc. This discussion can be linked to the wider literature on the use of learning analytics.

In the introduction (P8-9) and relevant sections of the results and discussion (P27-28), we have discussed opportunities for support, intervention, monitoring and improving online collaborative learning.

Robert A. Hanneman (Reviewer 2): MEED-D-16-00778

How the Study of Social Structure Can Guide Course Design and Predict Student's Performance in a Medical Course

This paper shows that certain aspects of the structural location of individual medical students in on-line discussions are correlated with course performance, and that a combination of structural factors may predict course performance with a degree of accuracy that makes it, potentially, useful to monitor and perhaps manipulate structural locations to improve performance. An interesting and relevant data set is used, and the literature cited on structural effects on performance, and on network analysis more generally are solid and useful.

There are, however, a number of areas in which the paper might be stronger.

As currently written, the paper may be attempting too much. On one hand, a goal is to demonstrate a causal linkage between structural location in networks and individual performance. The article, however, also seeks to explain many network concepts and how/why...
these factors might be connected to performance. The article also seeks to provide a survey of the general uses of network analysis for description, prediction (perhaps to guide intervention), simulation, and social-structural engineering. Because it attempts so much, it is not (to this reader) sufficient in meeting any one of the goals.

I would suggest that much of the material explaining specific network concepts (e.g. the many variations of centrality measures) be simply dropped - relying on available good reference works.

The possible use of network analysis to identify at-risk students, design curriculum, and intervene/experiment are mentioned, but not developed.

These ideas might have great value to the journal readers. They are not supported by data in the current experiment, but they might be discussed at greater length in the discussion/conclusion - so that readers might be stimulated to generate or monitor network data on their courses, and use it to improve teaching effectiveness.

There are a number of significant problems with the effort of the paper to effectively demonstrate that individual's position in the network matter in determining performance. At present, the paper (may) succeed in demonstrating a correlation between position and performance - an important caveat to this, though, is that the statistical significance testing method is not explained, and may be incorrect.

There are five main issues that need to be addressed in the effort to demonstrate a significant causal link between structural position in social networks and individual outcomes.

Endogeneity: it is plausible that a student's performance in the course is a cause of structural position, as well as the inverse. Students who perform very well at the beginning of the course (or who have a strong prior track record) may enjoy a "preferential attachment bias" (or "Matthew effect") such that they are likely to attract more social ties and become increasingly central as the network develops.

A dynamic simultaneous equation model of performance and outcomes would be called for to fully address this. Not possible in the current case, a control for performance at the beginning of the course (prior class standing, prior grades or test scores) might help. One could also see whether structural position at an earlier point in the development of the network was predictive of later position in the network.

Selection: Related to the above, actors with fixed attributes (e.g. gender) or varying scores on time-varying attributes (e.g. number of messages posted), may tend to form networks in different ways. Most notably, actors may differ (for both measured and unmeasured reasons) in their desire to seek ties, seek ties with central actors, or be seen as attractive targets of ties by others. These factors act as "network selection" biases. Notably, actors who are more likely to perform
well because of differences in talent, prior learning, or social skills are likely to both form bigger networks and perform better in this course - independent of any additional effect of network structural position.

Diffusion/social influence: The current analysis looks at the structural location of actors in their ego networks and the total network. In addition, though, the attributes of the alters to whom an ego is connected may matter. Suppose two students are each connected to five others. Student A is connected to five high performing students; student B is connected to five poorly performing students. The performance of one's alters may set expectations and enhance or limit learning resources for ego. The point is, the author's might want to consider adding attributes of each ego's alters as predictive factors.

Lack of theory/hypotheses: The article asserts that there are benefits to enhancing performance of structural position in networks. The article does a good job of citing some relevant literature. But the article does not do much to provide theory of the supposed processes by which such effects may occur. Social influence, social support, access to resources, ability to broker and compete effectively, etc. could be developed more to make a more compelling case for why we should care about structural position and hence use it in designing curriculum and intervention.

Failure to control: Most literatures on variation in individual performance in a wide variety of contexts suggests that variation in individual attributes matter. They probably matter indirectly, by influencing network selection. They probably also matter directly, regardless of the structural position of actors in networks. The current analysis does not control for likely attribute effects (e.g. age, gender, prior experience, appearance, general sociability, and prior academic performance - and likely other things). As such, the failure to provide controls makes the case for structural effects suspect.

Although some sentences in the original manuscript suggested that “network influences …… matter for performance above and beyond individual attributes factors”, our position is that network position or SNA parameters are not the only determinants of performance, or matter more than attributes, we think that they might be of benefit as performance indicators especially in collaborative settings.

We have removed all such instances and as to the reviewer suggestion “softened to only assert that network structural effects may provide additional insights.”

We have re-written the theoretical section on why network position and interactions matter to performance, added literature that tested the effect of network position on performance, and detailed their findings  page 8-9-10.
The Eigenvector centrality were used to study the alters and how this correlate or predict performance, it was one of the most important factors that correlated or predicted different performance measures and at all points of measurement.

More details and theoretical discussion were added in different section to explain how enhancing interactions, or social profile might affect performance, an example the first paragraph on page 8 first paragraph.

in the introduction (P8) and relevant sections of the results and discussion (P27-30), we have discussed opportunities for support, intervention, monitoring and improving online collaborative learning for a course, or an individual student.

Some particular details:

Purpose of the paper (currently page 11) needs to be put right up front.

Both the literature review and the discussion of networks need to focus on predicting, and potentially intervening, to improve student performance. At present, there is little theory or hypothesis explicit. There need to be compelling logical arguments (theories) that suggest why network position (and what exactly about it) and network influences (and how they work) matter for performance above and beyond individual attributes factors.

The title and first research question were re-phrased to clarify the purpose of the paper.

More attention was given to theoretical background of predicting performance and intervention, which resulted in re-writing large parts of p8, 9,10.

• We have expanded the section on the predictive potential of Social network analysis.

• We have re-written the last three pages of the introduction to discuss the literature on the subject of using SNA in learning analytics. We have also improved the methods section by adding more details to the SNA measures used in the study, description of the predictive algorithm and the statistical tests.

• We have expanded the analytics results by including data for early participation (correlation and prediction of midterm and final performance and logistic regression for predicting underachievers.). Reviewer 2 recommendation.
• We have also expanded the final performance results by including results of clinical performance, knowledge and comprehension (MCQ) and final grades. Reviewer 3 recommendation.

• These findings were further discussed in the discussion section.

We believe that the visual analysis is very important as an easy way to have a general view of a course, a discussion or a single student. Interpreting SNA parameters without the visualization might be misleading or incomplete.

We have removed the table of the two students and expanded the section of analytics.

7. 9-14. This claim is too broad. Most SNA perspectives argue that attributes and structural position are co-evolutionary and also somewhat independent. While fixed attributes may affect position in the network; dynamic attributes (e.g. behavior, attitudes) may be the result of position. The effects of fixed attributes (e.g. gender, age) may be both direct and indirect (by way of network selection).

That section was reworded as follows:

“Advocates of social network analysis emphasize the role of the social structure, one’s position in the community, and one’s relations and interactions as important factors that shape one’s learning and experiences, in addition to attributes as age, gender, and disposition [33]. Two types of SNA analysis are usually used, network visualization and analysis.”

7:42 among not between; node size proportional to "degree"; centrality has not been defined. Centrality should be linked to the current problem: why or how does being central in various ways affect performance?

“Among” was corrected, the section on centrality measures was re-structured, further explained and examples were given (P13-14) under the heading network analysis. We grouped importance of contribution into three main roles (Quantity of activity, Role in information relay and connectedness).

We have re-written the theoretical section on page 8 and 9 and detailed why network parameters might be considered as indicators of performance, or give be of potential benefit.

10: 18 "A network" not network. Weak discussion of degree distribution - needs precise definition and brief explanation of why one would care in the current problem context.
This short paragraph was dropped since degree distribution it was not used in the analysis.

10. Last paragraph of section identifies several additional important network properties, but gives only one word definitions and doesn't link these to the current problem.

Section 3 on simulation does not have a very clear definition of the processes and purposes of network simulation (simulations of action on the network? Or of the dynamics of the network?). Is simulation used in the current work? If not, perhaps drop this, too brief, section. If used, how is it applied to answer questions about variation in performance?

We made sure each centrality measure used is clearly explained with examples (P13-14)

Since simulation was not used in the analysis, this paragraph was dropped according to reviewer recommendation.

Section 4 on intervention is useful, but too brief. Several new implicit hypotheses are introduced here - suggesting things like poorly connected actors may be at higher risk of poor performance, and - if identified early - might have their networks re-engineered to reduce risk. But, the hypotheses must be made explicit. Perhaps recommendations for interventions in the teaching context should be in the discussion/conclusion, and based on the findings.

Intervention was covered in details in the introduction page 1st paragraph of page 8, and in the relevant section in the discussions especially p 27

8. p. 11: Can you give one or two concrete examples how SNA may help to inform interventions?

Three examples of how SNA may help to inform interventions were given on page 1st paragraph of page 8

11:20

12: 55 "Moodle" and "LMS" are not clear.

This phrase was re-worded to clarify the meaning
13:42 insufficient description of the graphics algorithm. Is this similar to 2D MDS on a valued matrix, with spring embedding? What does it maximize? What are its limitations for the current application?

We have improved the description of the algorithm, and added three references that fully explain the algorithm, its theoretical basis and comparison with others.

We kept the description simple, since the algorithm description is very technical and is not of interest to the wide audience of a medical education journal, however we have added three reference that offer detailed description of the algorithm, comparison to others and theoretical backgrounds, so that interested audience can refer to.

13 bottom 14 top. Not all of these metrics have been explained or mentioned in prior discussions. Hypotheses should guide which metrics are examined, and the metrics should be given formal definitions either here, or in earlier sections.

The centrality measures on importance was re-structured, further explained and examples were given (P13-14) under the heading network analysis. We removed the redundant centrality measures (Harmonic c and PageRank) since other measures did their function. For clarification, we grouped importance of contribution into three main roles (Quantity of activity, Role in information relay and connectedness).

The theoretical section on predicting performance and intervention was detailed on page 8-9.

We made sure each centrality measure used is clearly explained with examples (P13-14)

14: The discussion thread level is interesting. Why was it used, in addition to individual and whole-network? Perhaps thread should be treated as a replication or blocking factor? Did discussions occur sequentially? If so, they might show how earlier network patterns affect later developments.

Throughout the whole paper, there is a tension between explaining the network analysis and its relevance at the general level, and presenting the current case. Perhaps the writing could, in each section, first give the general discussion, and then present the current case example as an illustration. Or, one could do a hypothesis testing exercise on the current data, and then use it to draw general lessons in discussion/conclusion.

We have clearly explained this in the results and the discussion,
The importance of profiling an individual discussion is that it is the building unit of the course network, diagnosing gaps or pitfalls in a dysfunctional online group collaboration starts at the thread level and if intervention would take place, it usually happens at a discussion level.

Research has shown that teacher intervention in small group might be necessary in certain situations, such as a dominating students, or dysfunctional communications [11, 37]. Monitoring individual discussions by means of SNA can help inform instructors about when to intervene, in the example we demonstrated, the non-participatory discussion was a candidate for such intervention.

We have calculated all network parameters up to midterm to study how early network events might correlate to performance. Correlation as well as linear and logistic regression was calculated and presented.

"total volume” What data were actually analyzed? Presence/absence of a directed/undirected posting? Were the number of words counted and a valued network used? Not clear. Needs to be clarified, and a rationale given.

Total volume of interaction was an explanation to degree centrality, the sentence was re-written to clarify the meaning.

Since the instructor plays a role in structuring the interaction, it might be useful to examine the data both with, and without the moderator. If the goal is to generate peer-to-peer interaction, the instructor should probably be removed.

Yes, the goal was to generate peer-to-peer interaction, so the instructor was not included.

The assertion is made that structural position has greater effect on outcomes than do individual attributes. This is not tested or demonstrated in the current paper. It would be nice to have node attributes such as demographics, academic record and experience, attitudes, etc. If such controls are available they should be used in the analysis. If not available, the claims here must be softened to only assert that network structural effects may provide additional insights.

Don't know what automatic linear regression is. If it is an automated variable selection technique, it is highly suspect, unless the results are validated in additional samples. Better to use simpler and less automated modeling techniques.

As mentioned in the first comment, the assertion has been made clear that “to only assert that network structural effects may provide additional insights.”
The significance testing of both correlations and regression coefficients is not explained. This is not an independent replications sample, so conventional estimations of standard errors cannot be used. Permutation methods for both correlation and regression effects testing are essential.

We have re-done all analysis and used Kendall’s Tau-b for correlation and explained the significance levels where relevant,

We used both backward stepwise logistic regression and reported significance levels and other regression parameters,

In addition to Automatic Linear regression

We used automated model –although might not be favorable by some statisticians, for two main reasons; the first, to show the possibility of offering automated predictions which is the main purpose of analytics, to offer automatic predictions.

Second to rank predictors as a method to choose the most important predictors for stepwise regression,

SO, we satisfied both points of view (automatic predictions and “manual”).

Third, simple linear regression was not the best available option for our dataset that contained extreme values and simple regression does not handle such cases, we have explained this in the manuscript with a reference to literature that explains the advantages which we sum them here:

ALM was chosen over traditional regression models due to the following reasons

- Inability to handle outliers (extreme values that exerts an exaggerated effect on the model and results), since some students were either very active or very inactive (We had 6 predictors with outliers), the problem of extreme values needed to be dealt with. Using linear traditional regression would not solve this problem.

- Not possibility of doing the best subset method for variable selection.

- For variable selection, optimality statistics in multiple linear regression is limited to significance tests which is prone to Type I/II errors.

- No possibility to conduct ensemble methods to for better predictive models.

The present manuscript describes how social network analysis (SNA) can be used to uncover the nature of social learning processes within communities of learners as well as the impact of these uncovered processes on individual achievement. The study was run in an online surgery course for Medical students in which they engaged in clinical case discussions. Using a web-based online forum, students had the opportunity to exchange messages with each other and the instructor. Social network analysis was used to (a) visually represent the flow of the discussions, (b) to extract a variety of calculated network parameters and (c) to use these parameters as correlates and predictors for student achievement. The results showed that the visual representations that were created led to the identification of interesting discussion patterns both on an individual level as well as at the course level. Also, it appeared that some, but not all of the extracted parameters seem to be of importance (predictive value) for student achievement. While a considerable amount of variance in student grades could be explained by a conglomerate of all the extracted SNA parameters, it seemed that only information centrality and Eigenvector centrality were significant predictors of student grades. Additional correlational analyses yielded a couple of more relations between network metrics (e.g., in-degree) and centrality measures (e.g., Closeness centrality) with student grades. Overall, the study is innovative and demonstrates the potentials of using a still somewhat novel analytical technique to gain a better understanding of social learning processes that are occurring within communities of learners.

Thus, while I am generally positive about the manuscript, I believe that several aspects need to be worked out to recommend it for publication in Medical Education. More specifically, I have two major and a couple of minor concerns:

Major concerns:

1. My most important criticism refers to the question "What exactly do we know more after having read this manuscript?" While I do see the potentials of SNA to uncover processes that may not as easily be observable with other techniques, I am wondering whether this actually really is "big news". As the authors themselves describe, there is already plenty of research available that has used SNA for similar purposes (granted, not in Medical Education though) that has demonstrated the potentials of the technique. In large parts of their own empirical study, it seems that the already known potentials of SNA are just "demonstrated once again". Where it gets more interesting, though, is when the authors look at the correlations and predictive value of the different network parameters with and for student achievement. There, it becomes obvious that with SNA, the researcher is becoming capable to rather quickly (i.e. without cumbersome and time-consuming content analysis, for example) identify aspects of the social learning process that actually matter for learning. To me, this leads to recommendation that it would be good to shift a bit the focus of the analysis - away from "demonstrating the already known" and towards the impact that different aspects of the learning process that have been uncovered through SNA have on achievement or perhaps
further criterion variables the authors might have available. For example, I was wondering whether it would be possible to check whether the two discussion threads that are described on p. 18 have contributed equally or differently to student achievement. Perhaps it would also be possible to include a somewhat less general criterion variable than student grades (e.g., point scores for particular aspects of knowledge etc.)?

The analytics section was markedly expanded (p21-p22-p23-p24-p25-p26)

We have further expanded the analytics sections to include clinical performance, knowledge and final performance, we have also demonstrated how different predictors fare in different scenarios and further studied early participation up to the midterm.

Although SNA has been used before, our approach offered new insights regarding unstudied areas such as dynamic networks in the time-lapse video, the single thread discussions, and the information centrality plots, it is also according the reviewer the first in medical education and in the context of studying clinical case discussions.

We concur with the previous studies in their belief that context and course design matter. We believe that the significance of each indicator should be translated within the context of the time it was recorded, the setting of the online collaborative learning and the social structure of the course (which is best understood by the means of SNA visual analytics). In this study, the out-degree was an early indicator of good performance. As the course advanced, in-degree was more significant, as it represented an indirect vote of the quality of the contribution that attracted peer responses, in a course where the instructor was prominent, a prominent student is probably a distinguished one. The optimum usage of SNA in evaluating online collaborative learning should not separate centrality measure from visual analytics, but rather combine them to better understand the context and interpret the inferences of each indicator. And this was the main reason we included the first part of the results (visual analysis).

2. Related to this, I found one point in the authors' line of argumentation rather weak: Repeatedly across the manuscript, they argue that data that can be obtained from SNA are likely to be more closely connected to learning than "attributes such as age, gender, and disposition" (p. 7). I wonder whether comparing the effects of parameters that can be obtained via SNA to the effects of learning prerequisites on learning outcomes actually is (a) fair and (b) informative. Of course, what is the closest to learning outcomes will always be the processes that preceded these outcomes, not some very distal and general predispositions of variables such as gender or age. A fairer and more informative comparison would be one that looks at the predictive value of SNA parameters on the one hand and process indicators that may be derived from (admittedly more complex and cumbersome, but on the other hand
semantically more rich!) content analytical approaches, because both of them refer to the
same entity: the learning process. Having said this, I would recommend to make this
comparison a more central one, at least in the theoretical part and in the discussion. It might
even be feasible to use the available data to run such a content analysis and compare the
results of that to the results of the SNA. Yet, I do not see it as absolutely necessary to run
such a cumbersome analysis to make the point I just described. Thus, it should be enough to
address it in the theoretical part and the discussion section.

The instances that suggested that SNA is “more closely connected to learning than attributes
such as age, gender, and disposition” were re-worded, actually, our position is that data obtained
from SNA can add to our insights and are in no way the only or the determinants of performance.

The discussion of how SNA compares to the insights derived from content analysis were covered
in page 30

Minor concerns:

1. Large parts of the theoretical part and also of the results section read a bit like a
"commercial" for SNA. This impression is partially due to the fact that large parts of the
manuscript seem to treat SNA as some kind of end in itself, but is at least partially coming
from sentences that attribute "inherent power" to the technique (like it would be that once
you choose SNA, your job as a researcher is done). E.g., on p. 4: "Social network analysis
can analyze the social structure of a course..." - I would try to change such wordings into:
"By aid of social network analysis, researchers are empowered to analyze the social structure
of a course...". There are many instances in the manuscript that are similar. These should be
changed. (another example: p. 15: "The first sociogram... demonstrates the ability of SNA to
summarize all course interactions").

All such instances were re-worded “By using social network analysis, researchers can analyze
the social structure of a course” P4

“Using SNA to visualize learning networks, educators can have” P7

and elsewhere

2. In the background section, at several instances I found statements that were too strong for my
taste. E.g., on p.5: "It (TEL) also facilitates networked learning by means of computer-
supported collaborative learning - features that have been proven to positively and effectively
enhance learning". This sentence gives the impression that once you use TEL, everything
will be fine. In contrast, a lot of CSCL research has shown how important it is to design TEL
in a meaningful and purposeful way to really support learning. Thus, this statement should be revised. Also, it would be good to give a couple of references to CSCL research on scaffolding and scripting collaboration that illustrates this point (e.g., the work by the group around De Wever in Ghent or Fischer/Weinberger in Munich).

This sentence was reworded to be “features that have been demonstrated to positively enhance learning when coupled with properly designed resources”

The references were used in relevant sections, De Wever work on group collaboration and on content analysis; Fischer work on Script Theory of Guidance.

3. P.5: "Academically, it (collaborative learning) simulates skills of critical and higher-order thinking... and can be used as an alternative method for evaluation and feedback" - the first part of the sentence again seems to be too general to me (there is plenty of descriptions in the literature that show that collaborative learning can also become horribly wrong!); the second part I do not understand. How can collaborative learning become an alternative method for evaluation and feedback?

Both sentences were dropped and this part of the paragraph was re-worded

“Well implemented collaborative learning can facilitate knowledge construction, encourage involvement and motivation of learners in the learning process [3, 7-11].”

You seem to juxtapose LMSs and SNA by saying that LMSs offer "limited insight into studying interactions", while SNA offers more of such insight. I am wondering whether these two concepts are on the same page. I would rather see it like that: LMSs can give you the raw data that you need to be able to run analyses such SNA. So the former is more like a prerequisite for the latter. Or in other words: LMSs are not an analytical, but rather a pedagogical tool. SNA is an analytical tool.

All major LMS have built in dashboards that offer some insights of students’ activities, these dashboards usually capture quantitative measures of students’ activities which are inadequate for forums. We have reworded the sentence to clarify the meaning.

“the analytics dashboards built-in most major LMSs offer limited insight into studying interactions among students………. “

5. P. 7, line 49: Here you refer to the "importance" of a student for the discussion. It would be necessary to already know here how importance is determined (as you could also determine it in a content-analytical kind of way by rating how qualitatively high- or low-ranking that students' utterances are content-wise).
The section on importance (centrality measures) was re-structured, further explained and examples were given (P13-14) under the heading network analysis. We grouped importance of contribution into three main roles (Quantity of activity, Role in information relay and connectedness)

6. I did not completely understand all of the different SNA parameters that are introduced on p. 8 to 10. E.g.: "The harmonic centrality index is very close to closeness centrality index but better suited for disconnected graphs". This description still does not tell me exactly what "harmonic centrality index" actually is. The same a bit later when you talk about betweenness centrality: "It is similar to information centrality, which is another measure of a node in information flow and is an important measure of learners’ networks" - so what IS information centrality then exactly? I actually found the examples you gave on p. 10 helpful to get a more concrete picture of what at least some of the indicators mean and what they can be used for. I would thus recommend to add examples and research objectives to the descriptions of each of the indicators (instead of naming some of these only after you have introduced the indicators).

The centrality measures on importance was re-structured, further explained and examples were given (P13-14) under the heading network analysis. We removed the redundant centrality measures (Harmonic c and PageRank) since other measures did their function. For clarification, We grouped importance of contribution into three main roles (Quantity of activity, Role in information relay and connectedness)

7. p. 10: I would like to read more about the context of the study by Sie et al. - otherwise it is hard to understand sentences like this: "Their model showed that both powerful and less powerful innovators would benefit from the recommendation system..." - what exactly happened in the study? What was the task? How was behavior simulated? etc.

This paragraph was dropped according to 2nd reviewer, since it is not strongly connected to this study.

8. p. 11: Can you give one or two concrete examples how SNA may help to inform interventions?

P8 1st paragraph: three examples of intervention that can be informed by SNA, an isolated student, a non-participatory discussion thread and a dominating teacher.
9. p. 11, lines 45-57: This paragraph does not really fit into the flow of the argumentation, as the section overall talks about SNA and not about CSCL. Perhaps it would be good to add a short section "Context of the present study" right before the research questions section that informs the reader about the setting you used (that it was a CSCL setting, what tasks students had etc.). Then the line of argumentation would perhaps be smoother.

P 9. The paragraph was re-written as well as most of page 8-9.

P10 Context was further expanded and more details were presented about the CSCL,

10. p. 12: In the research questions section, the disbalance between the rather long description of processes that were uncovered by SNA and the rather short description of the relation between those indicators and student learning in the results section becomes especially obvious: Actually, you do not have a research question for all the results you present between p. 14 to 20! If you take the RQs seriously, you would actually need to label everything that is presented on p. 14 to 20 as "preliminary analyses" that are necessary to finally look at the relation of the indicators and student learning.

Thus, again my recommendation would be to shift the length of these parts of the results section towards a stronger balance. E.g., to answer the RQ2, I wonder whether it is really necessary to look at two students' networks to make the case. Perhaps this analysis could be replaced by one that provides more detail to the relations between SNA indicators and student learning outcomes.

The research question was re-worded and re-ordered. We believe that the visual analysis is very important as an easy way to have a general view of a course, a discussion or a single student.

We have removed the table of the two students and expanded the section of analytics.

11. p. 12: I was wondering how you can empirically answer RQ3. Perhaps it would be wise to take it out as a research question, but talk about its content as part of the discussion section?

The research question was re-worded and re-ordered. We believe that the visual analysis is very important as an easy way to have a general view of a course, a discussion or a single student. Interpreting the centrality measures can be best along the visual analytics that outline the context and the collaborators’ roles.

12. p. 13: "intended learning outcomes (of the course)" - what were they exactly?

Intended learning outcomes were added end of P10 and cont on p11 last paragraph.
13. p. 13: "Data were cleaned" - what does that mean?

A common problem for data extracted from a database is that some records might be corrupted, we explained it in the text p11

“The data were cleaned by removing corrupted records (two records were removed due to missing target of interaction),

14. p. 14: Can you quickly say what your goals are for each of the three levels of the visualization?

End of P 14 We have added the goals of visualization briefly and in details in each section namely P15 second paragraph, P 18 last paragraph,

15. p. 14: I am missing information in the results section on how student achievement was measured.

Those were added in details last paragraph of end of p11 and continued

16. p. 16, figures 3A and 3B: I believe there is a problem with the captions of figures 3A and 3B.

In the text, it says that figure 3A has to do without-going degree and that it demonstrates the information giving network, while the figure caption says "information-receiving network". Figure 3B should also read "in-going", instead of "out-going", I guess.

Regarding figure 3B, I wondered why none of the arrows point to somebody else than the instructor. The figure seems to imply that only the instructor received messages from students, but that virtually never one student received a message from somebody else. This is probably wrong though.

I also wondered whether your interpretation that the instructor dominated the discussion is valid. My interpretation of fig 3A and 3B would be that the instructor did receive many messages; but that he actually was not producing so many messages (as indicated by the larger node of S3 and S21 in fig. 3B.

The image was mislabeled and this created some confusion and this was corrected

The image shows some students receiving interactions, however not as frequent as the teacher, zooming in to S3 and S21 will show arrows pointing to them, as well as other students
The interpretation that the teacher received many messages and did not produce is in agreement with ours, we just highlighted that the teacher dominated the information-receiving network as follows:

P 17 - “the information-giving network was more collaborative, showed more active students and a moderate role for the instructor. The information-receiving network showed an exaggerated instructor role as he dominated the discussion over all the participants.”

17. p. 17, line 12: It should read “figure 4”, not "figure 3). That was corrected

18. p. 19: The network metrics seem to be interesting, but the information that is conveyed by the values is a bit hard to grasp. E.g., what does a "closeness centrality index" of 0.36 mean? Are there some sort of threshold values by which you could say that closeness centrality is "high" (just as there are conventions regarding for example how to interpret effect sizes or correlations as "weak", "average", or "strong")?

The scores are normalized so that 0 is the least and 1 is the highest, we have added this information to the table P20.

19. p. 21, lines 45-53: formulations such as "degree and out-degree were not significant" are problematic. More precise would be "degree and non-degree did not significantly correlate with student grades. Please check for more of such formulations (e.g. also on p. 27, line 13). Also: the correlation coefficients of degree and out-degree with student grades seem to be very close to significance, and the size of the correlations is also comparable. So, I would be more careful with making the implication that in-degree is much more strongly correlated with student achievement than the other two. If you would test whether the different correlations are statistically significantly different from each other, you would probably find that they are not.

These formulations were reworded in this section and throughout the manuscript, the correlation test was repeated using Kendall’s Tau-b, in-degree was also the significant factor,

We believe that context matters, we explained why in the manuscript
“what mattered was how students were able to establish their cognitive presence as judged by
their peers to be worth discussing [6, 17]. According to Chi et al. [62], interactions may be
considered beneficial only when they generate knowledge that is beyond the presented learning
materials and beyond peer’s contributions, which deserves peers replies and discussion.

20. p. 21, line 50: Nevertheless, I find the result that in-degree centrality is significantly
correlated with student grades, while the other two degree indicators are not, theoretically
interesting: This seems to imply that it is more important for my learning that I am addressed
often by others but rather that I contribute something to the discourse. In the interpretation of
this finding, you could refer a bit to recent learning activity models such as the ICAP model
by Chi and Wylie (2014).

Thank you for the reference, this is linked to the previous comment, in our situation, many
students read the posts of their peers, and they were mostly rewording them and offering no real
contributions, however, others offered real arguments, those were probably the better students,

According to ICAP model by Chi and Wylie (2014).

“interactions may be considered beneficial only when they generate knowledge that is beyond
the presented learning materials and beyond peer’s contributions, which deserves peers replies
and discussion”

22. p. 23, line 4: I did not quite get what the advantage of the automated linear regression is
compared to a multiple regression. If you use a multiple regression with a backwards
procedure, wouldn't you have the same effect as the one you have with the automated linear
regression (i.e. in step 1 you would include all SNA predictors, and then you would reduce
them from step to step by the one predictor that explains the least additional evidence)? Also:
Is it possible to present all R-squares and p-values for each SNA indicator?

ALM was chosen over traditional regression models due to the following reasons

- Inability to handle outliers (extreme values that exerts an exaggerated effect on the model
  and results), since some students were either very active or very inactive (We had 6
  predictors with outliers), the problem of extreme values needed to be dealt with. Using
  linear traditional regression would not solve this problem.

- Not possibility of doing the best subset method for variable selection.

- For variable selection, optimality statistics in multiple linear regression is limited to
  significance tests which is prone to Type I/II errors.
• No possibility to conduct ensemble methods to for better predictive models.


We used ALM with boosting (100 iterations in which it chooses a subsample to reach a more accurate model), ALM helped us rank the predictors and use the most important ones in a stepwise backward logistic regression.

The use of automated model –although might not be favorable by some statisticians, for two main reasons; the first, to show the possibility of offering automated predictions which is Analytics is about,

Second to rank predictors as a method to choose the most important predictors for stepwise regression,

SO, we satisfied both points of view. (automatic predictions and “manual”).

23. p. 23, line 11: "The most important predictors were information centrality 45.2%..." should be changed into something like: "Information centrality explained 45.2 % of the variance in student grades" (if this is true).

That was explained in the text p 22, here we quote it:

“Predictor importance is a method by which SPSS characterizes the importance of each predictor and refers to the residual sum of squares if the predictor was excluded from the model. The values are normalized so that the sum is 100%”

24. p. 23, line 14: "The fact that SNA was able to explain..." - I think you cannot say that SNA "is able to do anything". Rather, as said above, it empowers the researcher to do certain things. Similarly, on p. 27: The accuracy of the ALM prediction was 37.7%".

All instances were re-worded and changed.

25. Finally, although the overall language style is good, there are still a couple of errors across the manuscript. I would therefore encourage the authors to have a native speaker check the revision.

We have checked the English and scrutinized the manuscript for style and language.