Appendix for: Network analyses of student engagement with on-line textbook problems

Jesper Bruun

Department of Science Education, University of Copenhagen, ster Voldgade 3, Copenhagen K, DK-1350

Pia J. Ray, Linda Udby

Niels Bohr Institute, University of Copenhagen DK-2100

Abstract

This appendix accompanies the main paper: Network analyses of student engagement with on-line textbook problems. Appendix A shows the mean scores of each cluster on each principal component. Appendix B provides descriptions of Groups A-D. Appendix C provides examples of sessions with our descriptions.

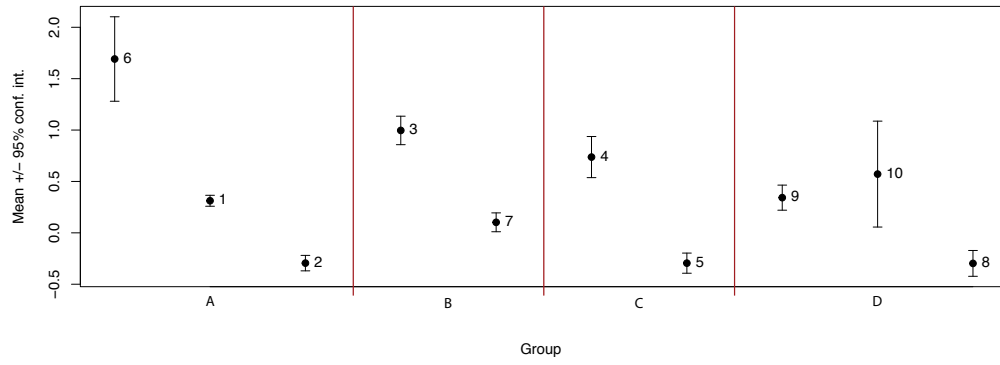# Appendix A. Groups A-D mean scores on components

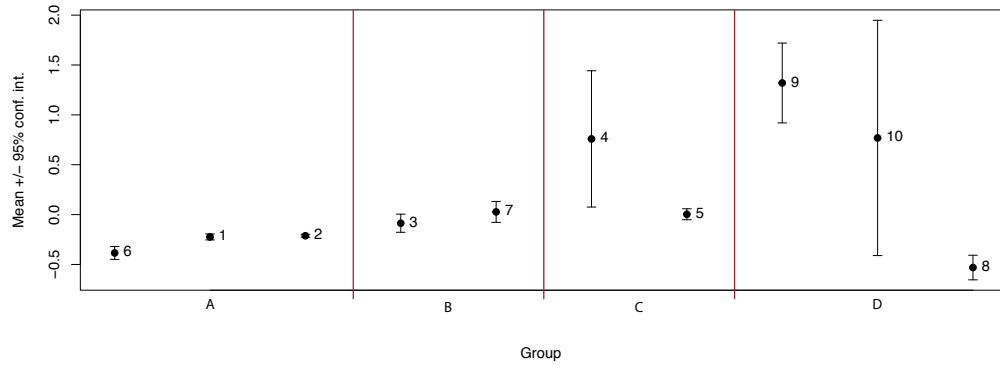Figure A.1: Mean scores on Component 1: Linear Length for clusters in groups.

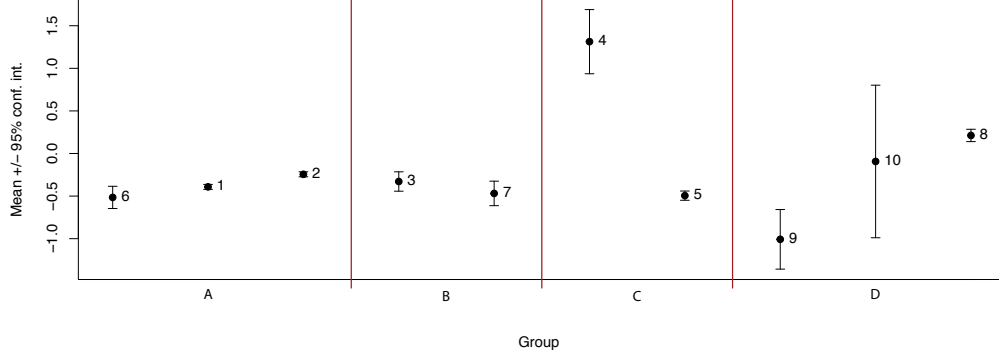Figure A.2: Mean scores on Component 2: Mutuality for clusters in groups.

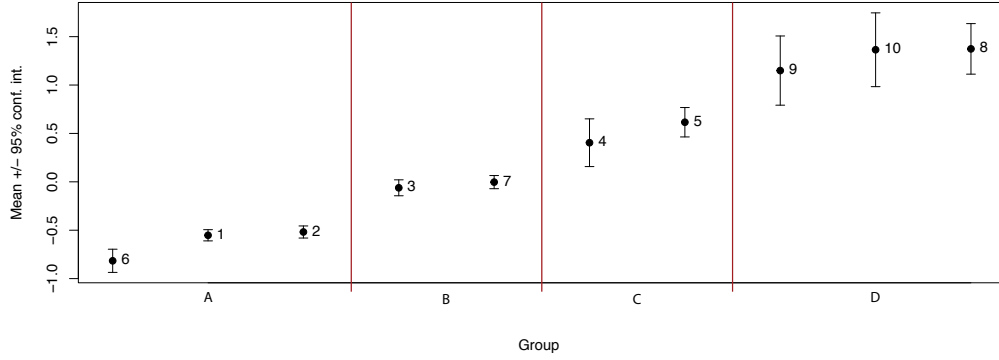Figure A.3: Mean scores on Component 3: Navigation for clusters in groups.

Figure A.4: Mean scores on Component 4: Complexity for clusters in groups.

Appendix B. Descriptions of Groups A-D

Here, we describe and interpret groups in detail.

Appendix B.1. Group A (Clusters 1,2, and 6) – the least complex group

<u>Description</u>: A total of 82 sessions are represented in this group. 75 of these sessions are full members of Group A, meaning that they are either fully part of clusters 1, 2, or 6 or their membership is shared between clusters 1, 2, and 6.

The clusters in Group A have significantly negative scores on the Complexity
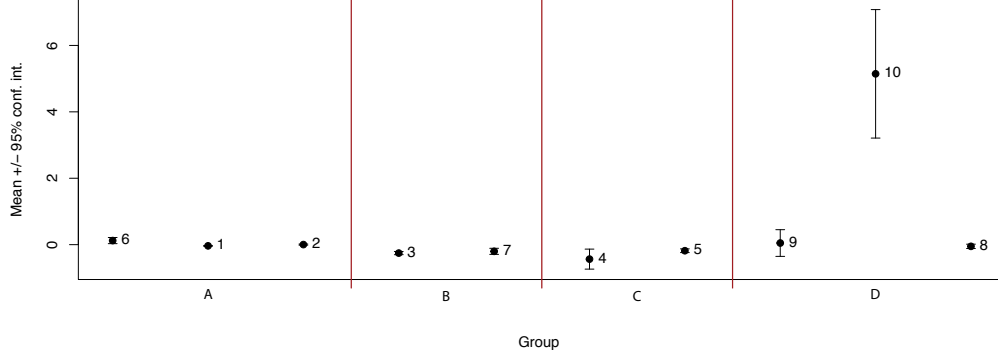
3

Figure A.5: Mean scores on Component 5: Erraticism for clusters in groups.

Component. Thus, Group A could be said to represent the least complex, or most linear, networks. Within this group, Linear Length separates the clusters: Cluster 6 scores highest of all clusters on Linear Length, Cluster 1 scores medium, and Cluster 2 scores low on Linear Length, so the linear length and average path length $\bar{l}$ varies within this group where session networks in Cluster 6 contain more nodes and links and session in Cluster 2 fewest. This can be confirmed by consulting relevant tables in METHOD X.

Appendix B.2. Group B (Clusters 3 and 7) – the medium complex group

Description: There is a total of 40 sessions in Group B, 23 of which are full members.
Group B score approximately 0 on the Complexity Component, which is significantly different than group A and C, i.e. sessions of group B has medium complexity.
Group B is also specifically separable on Target Entropy - higher than group A but lower than Group C. See [1]. Clusters 3 and 7 within group B is are separated by Linear Length. Cluster 3 scores significantly higher, meaning the linearity length parameter (as well as the average path length $\bar{l}$) is bigger in Cluster 3. our analyses confirms that Cluster 3 scores higher on every network measure that loads on Linear Length, while they score similarly on TE, S, and C [1].

4

Appendix B.3. Group C (Clusters 4 and 5) – the medium complex group

Description: Desription: 80 sessions are represented in this group with 55 being full members. Cluster 4 scores significantly higher than other clusters on the Navigation Component; Cluster 4 sessions have more V_{in}, V_{out}, and loop motifs when compared to other clusters. Cluster 5 scores low on Linear Length, approximately 0 on Mutuality, and significantly negative on Navigation. Interpretation:

Appendix B.4. Group D (Clusters 8, 9, and 10) – the most complex group

Description: There is a total of 62 sessions represented in this group, 41 of which are full members.
Mean scores on the Complexity Component in Group D are significantly positive and close to or above 1. Group D thus represents the most complex sessions of the dataset. Within group C, the clusters are further characterised on their mean scores on other components as described below.

- Cluster 8 scores low on Linear Length, negative (lowest) on Mutuality, and slightly positive on Navigation.

- Cluster 9 scores significantly positive on the Mutuality Component (but comparable to Clusters 4 and 10, which have large error-bars). I.e. these session networks show higher Mutuality than other sessions in Group C.

- Cluster 10 scores significantly positive on Erraticism but has large within group variation. This seems to be due to the large number of regulating/regulated mutual motifs.

Appendix C. Session examples in clusters

Appendix C.1. Example of sessions from group A (Clusters 2, 1 and 6)

Cluster 2. Figure C.6 shows a typical example of a session network from Cluster 2 with low Linear Length in the least complex group (A).
Description: The session is 6 hours and 7 minutes long but only has 5 nodes. The student starts at an exercise page and after ~ 1 minute he opens the solution to one of the questions. He spends close to minute presumably reading the solution and then goes to the overview page of all exercises where he after 20 seconds clicks on another exercise which is topically closely related

to the first. After roughly 6 hours where he presumably works with the second exercise (the particular exercise involves simulation and can thus be lengthy to solve) he shows a solution to a question in this exercise.

Interpretation: The session shows very little activity in the wiki-textbook probably due to activity elsewhere. The wiki-textbook is mostly used to look at formulation of exercises and check solutions after considerably working with them elsewhere.

The strategy of this learner seems to mainly use the wiki-textbook to read exercise formulations and verify solutions worked out elsewhere. There is very little use of interactive features of the wiki-textbook and no looking for info in text pages. This behavioral structure could be called Read.

to problem: neutron guide system 1

77.0

show solution

50.0

to page of all exercises

20.0

to problem: neutron guidesystem 2
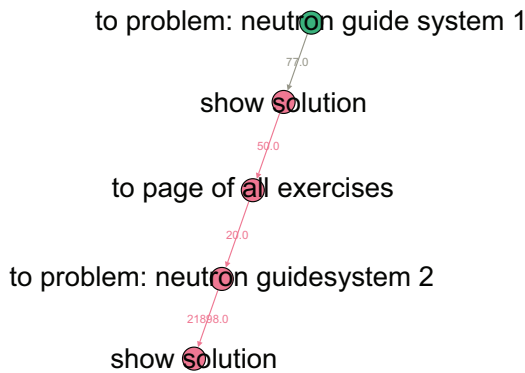
21898.0

show solution

Figure C.6: Session 1675 showing a typical example of a session network in Cluster 2.

Cluster 1. Figure C.7 shows a typical example of a session network in Cluster 1 which has medium Linear Length within the least complex group (A).

Description: The session has 10 nodes and takes 17 minutes. In this session the student starts at an exercise, looks at it for 15 sec, skips the hint and instead clicks on the solutions to the two questions of the exercise which he shows 29 sec apart. He then then goes to another exercise, opens the hint to the first question after half a minute and solution shortly after, but then spends 10 minutes (presumably working on the second question) before showing the solution to the second question. After a couple of minutes he then goes to a third exercise, opens the first solution after 19 sec but then spends more than two minutes (presumably working on the second question) before he shows the second solution.

Interpretation: This student seems impatient in the first question of exercises

to which he is very fast to show the solutions and prefer to spend more time working with the second questions before he shows the solutions. Maybe he thinks that the first questions are too easy? This behavioral structure could be called <u>Read-Selectively-Verify</u>.
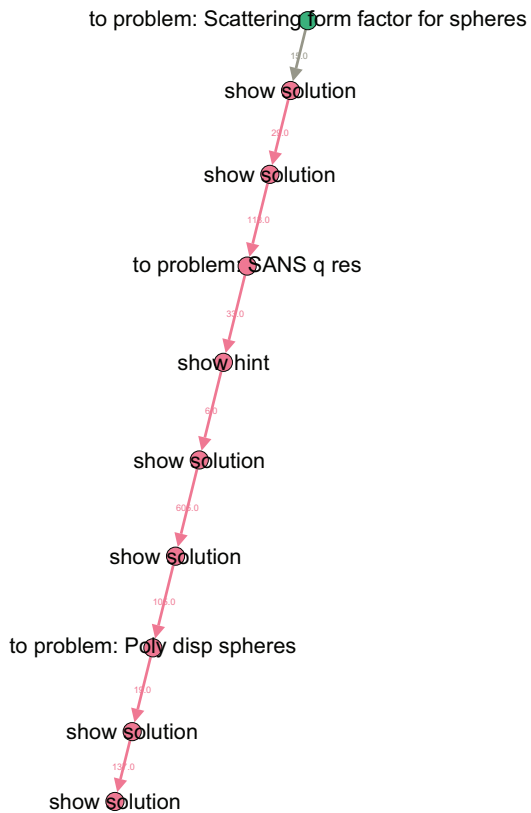

Figure C.7: Session 1046 showing a typical example of a session network in Cluster 1.

Cluster 6. Figure C.8 shows a typical example of a session network from Cluster 6, which has the largest Linear Length within the least complex group (A).

Description: The session has 24 nodes in close to 36 minutes. In this session the students starts from the main page, clicks after a few seconds on the main page of all exercises and then on a particular exercise after another few seconds. After ~ 3 minutes he shows the solution to the first question and after 27 seconds shows a hint to the next question. He then spends just over a minute presumably considering the hint until he opens the solution to the

question. He then spends another minute until he shows the solution to the next question. After a couple of minutes he clicks on another exercise and immediately shows the solution to the first question and after 18 seconds also shows the solution to the next question. After less than half a minute he clicks a third exercise and after a few (9-20) seconds opens solutions and hints to all of the questions consecutively (there was only one hint in this exercise). After 23 minutes he goes to a fourth exercise (via an exercise overview page), quickly (14 sec) after shows the first hint, goes to a relevant textbook page, quickly performs an action (search?) and returns to the exercise to show the solution after about a minute. *Interpretation*: At first glance this student seems to be very active. However after the first exercise the student seems to only glance at the questions, hints and solutions consecutively in a series of exercises and select out only a few specific questions in various exercises to spend around a minute on before opening the solutions. Only in one instance does he spend a longer time before performing the next action (opening another exercise) and it is more likely that he took a coffee break than worked with the previous exercise (and all the open solutions). Only in one instance out of the long consecutive "show everything quickly"-session does he seems to look for information and consider the question again for a minute before opening the solution.

This behavior could be explained by a student who printed out all information regarding the problems in the curriculum in order to solve them at another time, but in that case he would probably have spent less time between consecutive clicks. A more likely explanation is that this student shows a surface learning approach since he rarely spends more than a minute working with any question. He may verify the correct solutions without attempting a full survey-explore-plan-implement cycle of his own. This behavioral structure could be called Read-Verify.

Appendix C.2. Example of sessions from group B (Clusters 3 and 7)

Cluster 3. Figure C.9 shows a typical example of a session network from Cluster 3, which scores medium on Complexity and high on Linear Length within that group.

Description: The session total time is 34 minutes with 17 nodes. We see that the student starts on a simulation project page, shows the text of a simulation problem and then goes to a related problem where he quickly opens the solution to all the questions. After half a minute he then opens the first question in the simulation problem, then the corresponding hint and
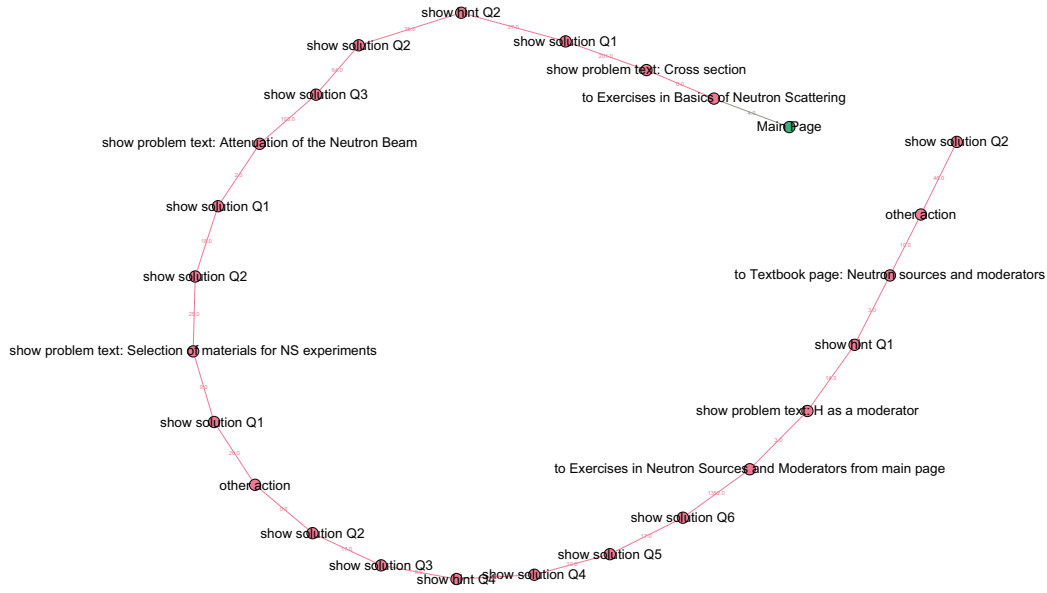
Figure C.8: Session 618 showing a typical example of a session network in Cluster 6.

following question of the simulation problem. He spends a minute, before reading the third question and another before opening the fourth and fifth which he then closes again after 2-4 minutes respectively to return to the first part of the project where he shows the hint to the second question after 4.5 minutes. After 20 seconds of revisiting the hint he then navigates several times to a related (simulation) problem where he shows the same solution each time he visits. The first time he shows the solution immediately, next he spends 12 minutes before clicking the solution, the third time he waits 48 seconds and the last time he shows the solution is after close to 9 minutes.

Interpretation: Since there are no solutions to the simulation project the student is trying to perform, he is in the beginning of the session reviewing the solutions to the related theoretical problem. He then returns to the project and presumably runs the simulation program in the background while comparing his results to the solutions of a related simulation problem in the wiki. This behavioral structure could be called Embedded-Read-Verify.

Cluster 7. Figure C.10 shows a typical example of a session in Cluster 7, which scores medium on Complexity and low on Linear Length within that group.

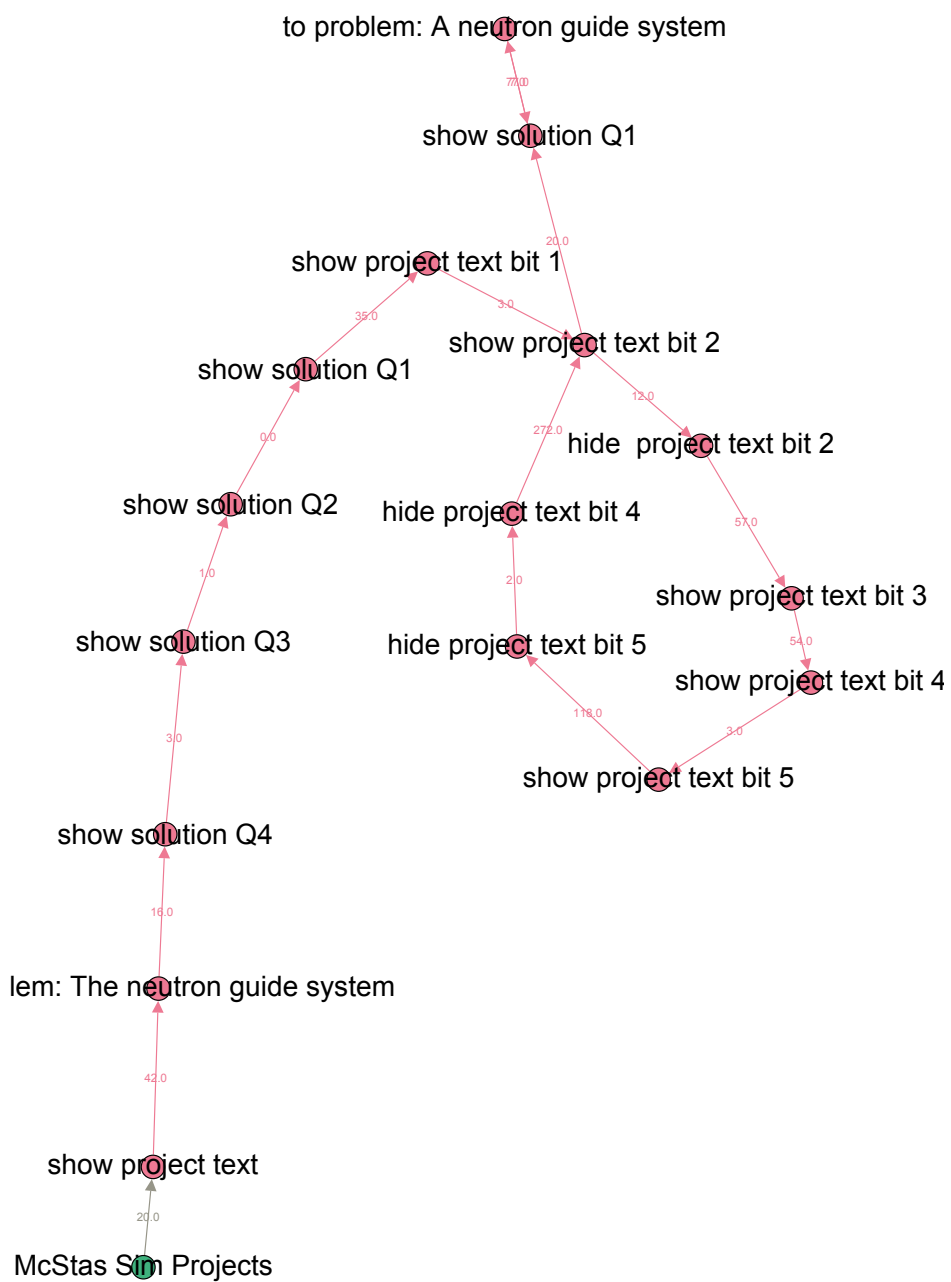*Description*: This session has 8 nodes distributed in 49 minutes.

9

Figure C.9: Session 813 showing a typical example of a session network in Cluster 3.

The student first spends 6 minutes on a specific problem before opening the first hint and then closing it quickly again to immediately open the solution of the first question which he only glances at (14 sec) before opening the solution to the third and last question. He also only glances at this before going to a second problem where he spends considerable time (19 minutes) before showing the solution to the first question, then spending 2 minutes before clicking on a part of the solution containing a figure which illustrates the solution and after 20 minutes clicking on the figure which gives a more detailed view of the contents.

Interpretation: The two problems with which the student works are topically related, the second one being a more complex case than the first. In the first question of the first problems the student is to derive a function which he then uses considerable time working on before opening the first hint, and probably sees that he has no further use of the hint so proceeds fast to verifying his solution. In the second and questions of the first problem he probably realizes that he needs a mathematical plotting program to answer the questions and quickly proceeds to the second problem. In the second problem he also needs to derive a mathematical expression which he uses 19 minutes to work on before verifying his solution. He then clicks on part of the solution containing the output from a mathematical plotting program and presumably implements his own function in a plotting program before verifying his results against the ones in the figure upon closer inspection. In this session the student the student works at length only with the questions not requiring the use of an external tool. He previews the solutions to questions which he needs to use an external tool in order to answer. Only in the end of the session does he implement the external tool and verifies only the most complex result. This behavioral structure could be called Embedded-Selectively-Verify.
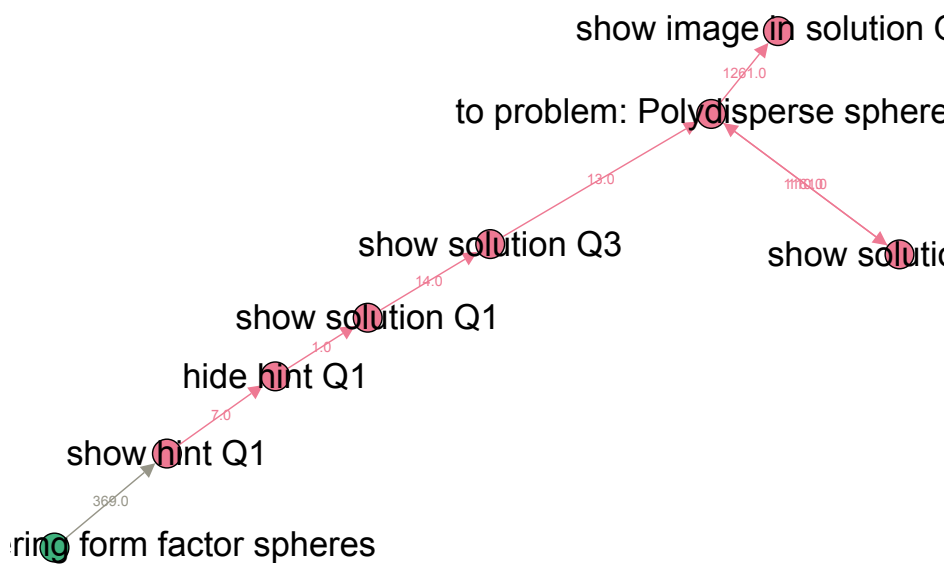
show image in solution

to problem: Polydisperse sphere

show solution Q3

show solution Q1

hide hint Q1

show hint Q1

ring form factor spheres

Figure C.10: Session 1826 showing a typical example of a session network in Cluster 7.

Appendix C.3. Example of sessions from group C (clusters 4 and 5)

Figure C.11 shows a typical example of a session network in Cluster 4 which scores high on Complexity, Linear Length, and Navigation.

Description: This session takes 42 minutes and has 16 nodes. The student navigates a bit on the main problem page and then after roughly half a minute clicks on a specific problem for which he after a few seconds shows the solution to first question and after about a minute shows the solution to the second question. After close to two minutes he navigates on the same page and then quickly thereafter opens the solution to the third question. After half a minute he navigates a bit on the same page and then after almost two minutes opens the solution to the next question. After more than 3 minutes he navigates to another problem and after spending a bit more than a minute he shows the solution to the first question but hides it again after glancing at it for 7 seconds. After spending a minute he then navigates to some related text pages where he spends 21 minutes before returning to the description of the problem, navigating a bit on the page for a minute. After close to 6 minutes he shows the solution to the second problem and spends a little more than one minute before proceeding to a third problem which he spends one minute reading before the session ends.

Interpretation: The student mostly spends at least a few minutes before opening solutions. It seems that in one case he realizes that his answer is wrong after glancing at the solution (quickly hiding it again) and therefore navigates to read information pages for a considerable amount of time before returning to the problem. He then probably works on the second question for some minutes before showing the solution. The student uses the solutions for verification of his solutions only after spending time on each question and searches for information in cases where his solutions are possibly incorrect. He thus takes advantage of the self-verification and easy access to related learning material features of the wiki-textbook. This behavioral structure could be called Read-Verify-Explore.

Cluster 5. Figure C.12 shows a typical example of a session network in Cluster 5, which scores high on Complexity and low on Linear Length and Navigation.

Description: This session takes close to 7 minutes and has 4 nodes. The student navigates first to an overview page of a particular subset of problems and after 12 seconds selects a particular problem. He spends 6 minutes before showing the first hint which he glances for 4 seconds before hiding it
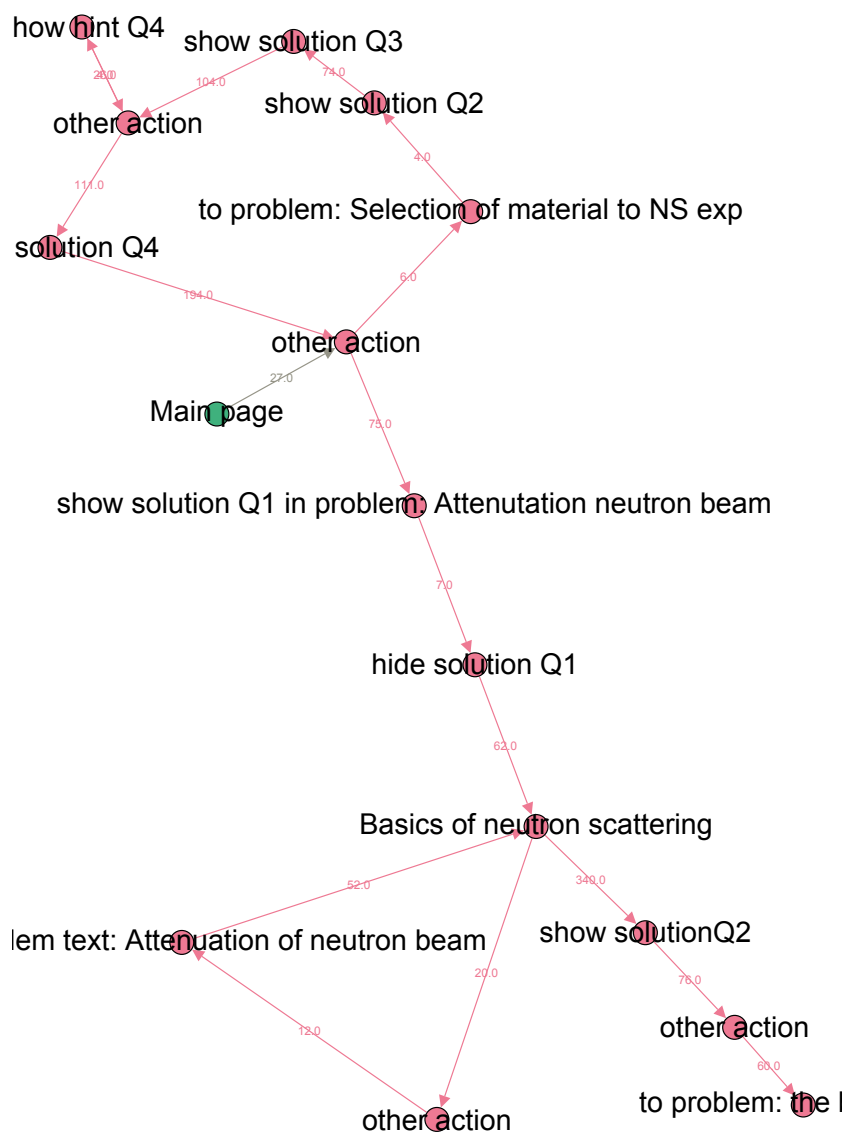
Figure C.11: Session 1591 showing a typical example of a session network in Cluster 4.

to Exercises in SANS

to Scattering form factor spheres
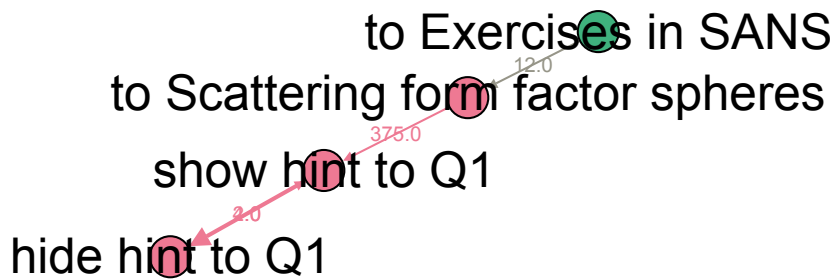
show hint to Q1

hide hint to Q1

12.0

375.0

3.0

Figure C.12: Session 118 showing a typical example of a session network in Cluster 5.

again. But after a few seconds he opens the solution once more to glance at it for a few seconds once more and then closes it again.

Interpretation In this session the student only works with one particular question which he first tries to solve and then several times glances at the solution (show-hide). He does however not work more with his solution between consecutive show-hides so probably he is uncertain about the first verification. This behavioral structure could be called Read-Peak.

Appendix C.4. Example of sessions from group C (clusters 8,9, and 10)

Cluster 8. Figure C.13 shows a typical example of a session network in cluster 8, which scores high on Complexity and Navigation, and low on Linear Length within that group.

Description: The session has 9 nodes and lasts for close to 8 minutes. The students starts the session by logging in, navigating to a page with text learning material but spends only about half a minute there in total before proceeding to the overview page of problems where he spends around 3 minutes before proceeding to first one textbook page and then quickly thereafter to another where he spends a couple of minutes before he navigates back to the overview page of problems. He then selects a specific problem and opens the solution after 41 seconds after which the session ends. *Interpretation*: In this session the student seems to guess about relevant background material for the problem he wants to do in advance of actually engaging with the problem. But after then looking quickly at the problem he may have realized that he did not find the right background information and simply shows the solution. This behavioral structure could be called Explore.

Cluster 9. Figure C.14 shows a typical example of a session network in Cluster 9, which scores high on Complexity and Mutuality and low on Naviga-
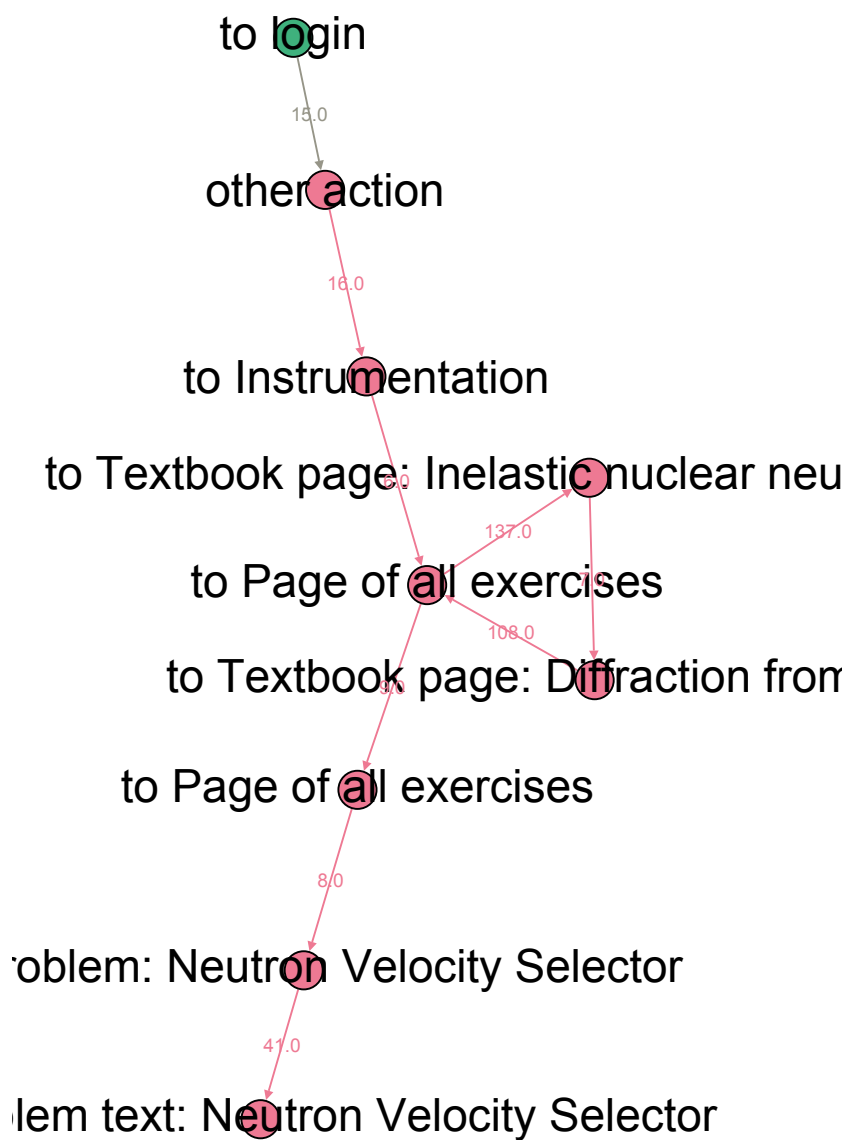
to login

15.0

other action

16.0

to Instrumentation

to Textbook page: Inelastic nuclear neu

137.0

to Page of all exercises

108.0

to Textbook page: Diffraction from

to Page of all exercises

8.0

oblem: Neutron Velocity Selector

41.0

lem text: Neutron Velocity Selector

Figure C.13: Session 1924 showing a typical example of a session network in Cluster 8.

16

tion. *Description*: The session lasts 332 minutes (close to 6 hours) and has 10 nodes.

The students starts at a textbook page with theoretical background for using particular simulation software and after 16 seconds he navigates to the overview of related problems. After half a minute he goes to the page of all problems in order to after another half minute go to a (complex) simulation project page where he shows the first question text. After 6 seconds he hides the text and then shows and hides the text quickly once more. After 12 seconds he then navigates to a simple simulation problem which is highly relevant for the first question of the more complex simulation project. He clicks on a headline and some text in the simple simulation problem during the next 45 minutes but presumably performs the required simulation to answer the question in another program meanwhile. Only after 45 minutes does he show the solution the first question in the simple simulation problem (there is no hint in this case). During the next 15 minutes he seems to be investigating the solution and perhaps trying to solve/simulate the second question (clicking various places on the simple simulation problem page without closing solution to Q1, each click is at least 3 minutes apart). When the 15 minutes have passed he opens the solution to the second question (there is also no hint to this question). After 2 minutes where he probably inspects the solution and click various places on the page he navigates to the overview page of all problems where the session ends. *Interpretation*: In this session the student seems to go directly from the background info to the relevant simulation project, then realizes that he needs to start with something more simple and then spend substantial time solving a simpler but highly relevant simulation problem without peeking on the solutions before working through the simulation himself. When he actually does open the solutions he appears to inspect them carefully, clicking various places in the page with open solutions several times. This behavioral structure could be called Integrated-Interactive.

Cluster 10. Figure C.15 shows a typical example of a session network in Cluster 10, which scores high on Complexity and Trial & Error, but shows high variability on the other components. *Description*: The session lasts 55 minutes and has 15 nodes.

The students starts on a particular problem page which he seems to inspect with 13 clicks on the page (headlines and paragraphs) distributed over 16 minutes in total before he shows the solution to the first question (there is
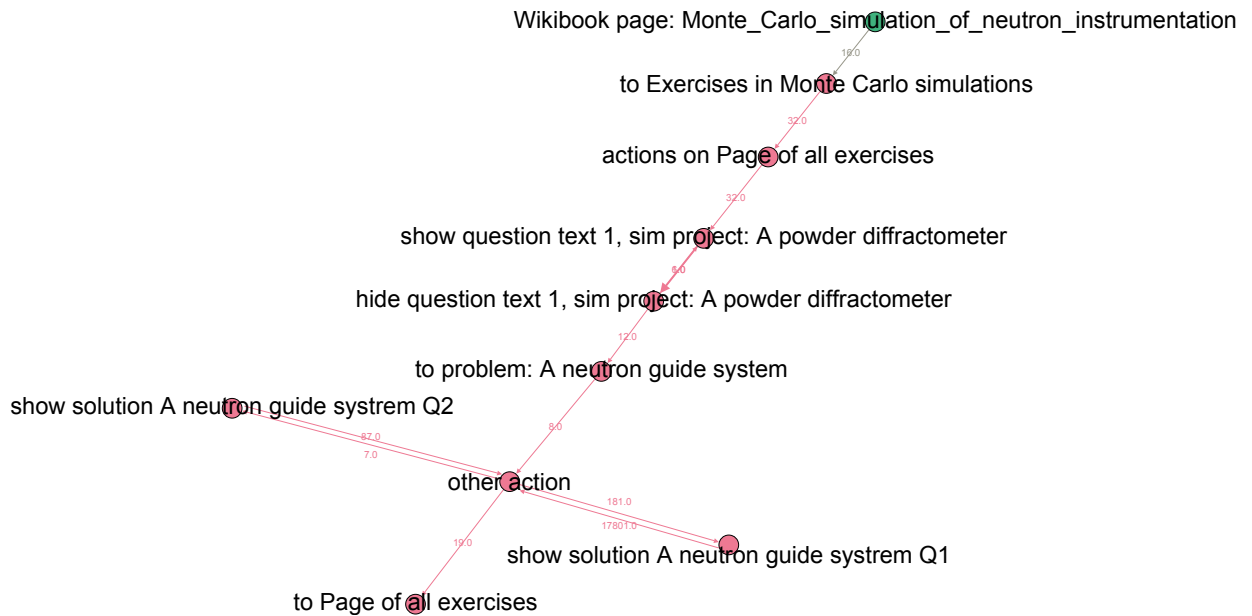
Figure C.14: Session 1649 showing a typical example of a session network in Cluster 9.

no hint). After glancing at the solution for 17 seconds he opens a related problem (which is posted in the solution). After a few seconds he closes the solution to the first question in the first problem and continues to work on the second problem, clicking a lot of times on various places in the page over the next minute. He the returns to the solution of the first problem which he glances at before closing it again to return to the second problem page where he after half minute opens the solution to the sixth question, inspecting it for 16 seconds before hiding it again and then quickly shows and hides the solution to the next question. Over the next 1.5 minutes he inspects the page of the second problem (clicks 4 times on paragraphs) before yet again returning to the first problem page which he inspects with 9 clicks over roughly a minute before yet again inspecting the second problem page with 28 clicks in roughly a minute after which he shows the solution to the third question, but hides it within seconds and then the solution to question 4 which he hides just as quickly. He then returns for a third time to the first problem, this time spending nine minutes inspecting the page in four clicks, before going a fourth time to the second problem where the inspects the page in some clicks and then after 2 minutes and 20 seconds opens again the

solution to the fourth question but hides it after a few seconds presumably pondering on this solution for 4 minutes before he inspects this page some more in 7 rapid clicks and then after 43 seconds returning to the first problem for the fourth and final time. Over the next 14.5 minutes he inspects the page, most click are only seconds apart but 2 of them are several minutes apart where he presumably is trying to derive the solution to the original question in the first problem for which he then shows the solution once more but hides again after a few seconds. *Interpretation*: The student seems to be comparing the content of one problems to the description and solutions of a second problem in the attempt to find a solution to the questions in the first problem. He opens some of the solutions (questions 6 and 7) in the second problems probably because there is one particular keyword figuring both in the first problem and in questions 6 and 7 of the second problem. But after revisiting the first problem he appears to not have found the help he needed and return again to the second problem to open some of the other solutions (questions 3 and 4). He may have got an idea on how to solve the first problem from these solution because he now spends nine minutes at the fist problem page, then he revisits the solution to q4 of the second problem which he uses some minutes to try and solve before revisiting (maybe verifying) the solution. He probably now considers the context of his result by inspecting the page of the second problem because after 4 minutes he inspects this page some more before finally returning to the first problem and presumably trying to solve it but also inspecting the page. After close to a quarter of an hour here at the first problem page he shows once more the solution but hides it again after only 2 seconds, which may be too fast for a verification of the result and could indicate he is rather giving up.

This learner is very persistent in trying to find information that could help him solve one particular problem which is formulated rather openly. He tries to find solutions to relevant problems and probably tries similar approaches to solving his problem. It is not however completely clear if he succeeds and he doesn't seem to spend an adequate time verifying his attempts (hides the solutions quickly). He makes the same steps several times indicating a trial-error approach to problem solving, possibly unsuccessful (in this case). This behavioral structure could be called Erratic-Interactive.

References

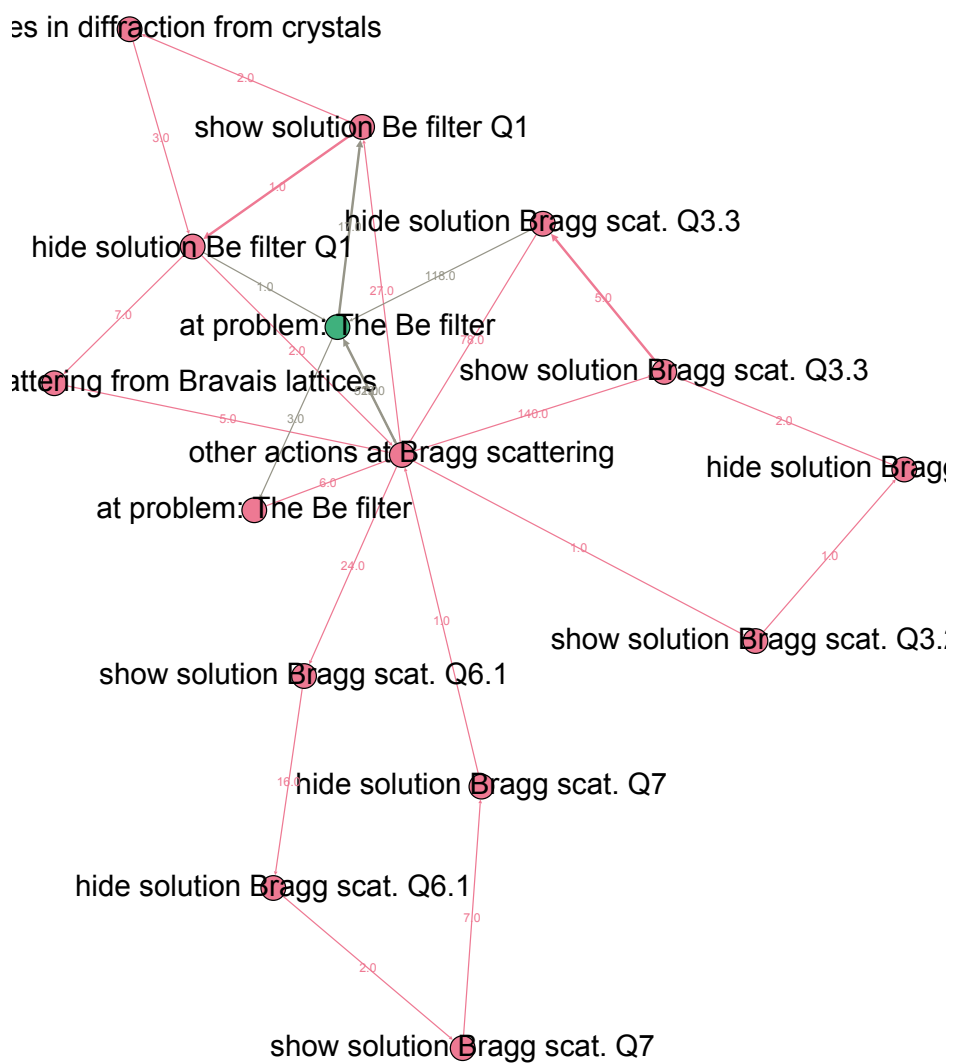[1] Jesper Bruun. Network analyses of computer log files, April 2024.

Figure C.15: Session 1938 showing a typical example of a session network in Cluster 10.