Network analysis of students’ use of representations in problem solving

Daryl McPadden\textsuperscript{a} and Eric Brewe\textsuperscript{a,\textbeta,\textgamma}  
Florida International University, 11200 SW 8th St., Miami, FL 33199  
\textsuperscript{a} Department of Physics, \textsuperscript{\beta} Department of Teaching and Learning, \textsuperscript{\gamma} STEM Transformation Institute

\textbf{Abstract.} We present the preliminary results of a study on student use of representations in problem solving within the Modeling Instruction Physics 2 course (MI-Phys2), which covers introductory electricity and magnetism (E&M). Representational competence is a critical skill needed for students to develop a sophisticated understanding of college science topics and to succeed in their science courses. In this study, approximately 70 students from the MI-Phys2 course were given a survey of 25 physics problem statements both pre- and post-instruction, over both Newtonian mechanics (NM) and E&M. For each problem statement, students were asked to select one or more representations they would use in that given situation. We analyze the survey results through network analysis (NA) in order to identify which students selected similar representations. We also compare the student networks for those students who had previously taken the Modeling Instruction Physics 1 course (MI-Phys1) and those students who had taken an alternative Physics 1 course.

\textbf{PACS:} 01.40.-d, 89.65.-s, 01.40.G-

\section{I. INTRODUCTION}

Communication of ideas and concepts inherently depends on the use of representations, frequently through words or pictures. Science is a particularly representation-rich subject area, with graphs, equations, verbal descriptions, diagrams, and pictures often used together to describe and understand a particular phenomenon. Scientists are not only competent in each individual representation but often must translate and integrate between them, making a coherent, varied model of a given situation [1–4]. As instructors, we want to develop the same experience in our students. Therefore, we aver that representational competence is a critical skill needed for students to develop a sophisticated understanding of college science topics and to succeed in their science courses and future careers [4–7].

University Modeling Instruction (MI) has been specifically designed to focus on representation use in introductory physics [5] and has been shown to be successful by multiple measures, including FCI gains, CLASS gains, and odds of success [8–12]. MI is a reformed curriculum and pedagogy where lecture, lab, and recitation are combined in a studio format that encourages students’ participation and agency through group activities and guided classroom discussions [5, 13]. Students are responsible for developing, testing, evaluating, and revising models, which are defined as a coordinated set of representations that can be used to explain, predict and describe a particular class of phenomena [2, 13, 14]. Brewe outlined that representations and their coordination should be the primary focus for the first half of the modeling cycle, with explicit discussions about how to use the representations and the information that they contain [5].

The focus on representations is a common theme throughout both the Modeling Instruction Physics 1 course (MI-Phys1), covering Newtonian mechanics (NM), and the Modeling Instruction Physics 2 course (MI-Phys2), covering electricity and magnetism (E&M). At Florida International University (FIU), there are many ways to meet the Physics 1 pre-requisite class for MI-Phys2: lecture/lab courses, courses based on an Investigative Science Learning Environment (ISLE) curriculum[15], and the MI course; in addition to transfer and advanced placement credit. Therefore, it is unlikely that every student in MI-Phys2 has taken MI-Phys1.

This leads us to the following research question. How do students’ Physics 1 courses impact their representation use across both Physics 1 and Physics 2 contexts?

\section{II. DATA COLLECTION}

In order to address this research question, we designed a survey of 25 physics problem statements that cover both NM and E&M topics, which was given both pre- and post-instruction in the Spring 2015 MI-Phys2 course. We will refer to this survey as the Problem Solving and Representation Use Survey (PSRUS). For each problem statement, students were given a list of 16 possible representations (e.g., graphs, equations, motion maps, word descriptions, field lines) and asked which representations they would use in that given situation. Students were not required to actually solve the problem, but simply indicate what representation(s) they would use to do so. Students could respond with anywhere from 1 to 16 options (including an “other” option). For example, one survey question states, “An asteroid falls toward a planet. What path does the asteroid take?” Possible relevant representations could then include a position vs time graph, gravitational field or potential lines, a picture, the gravitational force equation, and/or a force diagram.

Since there are no “right” answers and the number of representations varied by each question and student, we decided to use network analysis (NA) to determine how similar students responses were on the PSRUS. We built a network based on the representational choices of the students, where students who responded with the same representations were linked together. For example, if Student A and Student B both responded that they would use graphs on Question 1, they were marked as...
connected with a weight of 1, where the weight (given by \( w \)) represents the number of shared representations between Students A and B. If students C and D said that they would use equations, field lines and force diagrams, they would be marked as connected with a weight of 3. (See Figure 1.)

In this way, we created a student network for both the pre- and post- PSRUS data that includes student responses from all 25 questions. For the pre-student network, there were 69 students (i.e., nodes) and 2346 edges (i.e., connections between students). For the post-student network, there were 60 nodes and 1770 edges. In both of the student networks, the edges are undirected and weighted by the number of shared representations.

Of the 69 students who took the pre-PSRUS, there were 36 students who had taken MI-Phys1, leaving 33 students who had taken one of the other options. Out of the 60 students who had taken the post-PSRUS, there were 32 students who had come from MI-Phys1 and 28 students coming from the other courses. Thus, for both PSRUS surveys, there were about equal numbers of students who had taken MI-Phys1 and those who did not.

III. ANALYSIS METHODS

A. Inbreeding homophily index

In NA, homophily is a measure of how connected nodes with similar characteristics are, epitomizing the saying “birds of a feather flock together” [16, 17]. To address the research question, we split the students into two groups - students who took MI-Phys1 (Group M) and students who took another course for Physics 1 (Group O). The homophily index allows us to determine how representationally similar the two groups are based on the number of edges the nodes have within their group and between groups.

In order to quantify homophily, first suppose there are \( K \) groups of nodes in a network with a total of \( N \) nodes. For group \( i \) with \( N_i \) nodes, we define the relative fraction of nodes in group \( i \) as \( f_i = \frac{N_i}{N} \). Then, let \( s_i \) be the average number of edges between nodes within the same group, and let \( d_i \) be the average number of edges between nodes from different groups. First, we use Welch’s \( t \)-test to determine if the \( s_i \) and \( d_i \) means are statistically different. We then calculate the homophily index, \( H_i \), for a given group:

\[
H_i = \frac{s_i}{s_i + d_i}.
\]

(1)

From the homophily index, we then calculate the Inbreeding Homophily Index (IHI), which normalizes the homophily index based on the relative fraction of the group. The IHI is given by

\[
IHI_i = \frac{H_i - f_i}{1 - f_i}.
\]

(2)

If the \( IHI_i > 0 \), then group \( i \) is said to have inbreeding homophily, meaning members of group \( i \) are more likely to be connected to other members of group \( i \). An \( IHI_i = 1 \) indicates perfect homophily (all edges are among similar nodes). If the \( IHI < 0 \), then group \( i \) is said to have inbreeding heterophily, meaning the members of group \( i \) are less like to be connected to members of group \( i \). An \( IHI_i = -1 \) indicates perfect heterophily (all edges are between members of different groups). If the \( IHI_i = 0 \), there is no preference exhibited by the members of group \( i \) [16].

B. Walk-trap community detection

Using the walk-trap community detection algorithm, we can also determine the number of communities that are present in the PSRUS pre- and post- networks. The walk-trap algorithm is based on the idea that if a random path was chosen along the network structure, you
TABLE I. This table shows the results of t-tests between the average number of same edges and the average number of different edges for the student who took MI-Phys1 (Group M) and those who did not (Group O). All tests show that the number of same edges and the number of different edges were statistically different (p << 0.001).

<table>
<thead>
<tr>
<th>Group Pre/Post Deg. of Freedom</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>M Pre</td>
<td>50.64</td>
<td>5.65</td>
</tr>
<tr>
<td>M Post</td>
<td>52.65</td>
<td>6.49</td>
</tr>
<tr>
<td>O Pre</td>
<td>51.26</td>
<td>4.30</td>
</tr>
<tr>
<td>O Post</td>
<td>44.50</td>
<td>4.82</td>
</tr>
</tbody>
</table>

would be more likely to stay within a community since the number of edges inside a community should be higher than the number of edges between communities [18].

The walk-trap algorithm determines the final number of communities based on which grouping of nodes maximizes modularity [18]. Modularity of a network is a measure of how effectively the network splits into communities, based on the density of edges in a community compared to the number of edges. A higher network modularity indicates that there is “good” partition of the network into communities, with the community edge density being greater than that of a random network [19].

Once the communities are determined (as Community 1, Community 2, and so on), we label the M group as 1s and the O Group as 2s (chosen to maximize the correlation) and use general linear regression to determine the correlation between predicted communities and the actual M/O Groups.

IV. RESULTS AND DISCUSSION

A. Inbreeding homophily index

As explained in Section IIIA, we first used t-tests on the number of edges between similar nodes and the number of edges between different nodes for both the M and O Groups. The results of both the pre- and post-network t-tests are shown in Table I. All of the t-values were highly significant (p << 0.001), which means that the number of similar edges and the number of different edges were statistically different for both groups in pre- and post-networks.

The IHI was then calculated using equations (1) and (2). For the M Group, the pre-IHI was 0.238, which was reduced to 0.120 for the post-IHI. For the O Group, the pre-IHI was -0.128 and -0.109 for the post-IHI. Looking at the pre-IHI for the M Group, the positive value indicates that there is inbreeding homophily among the students who took MI-Phys1, meaning there is a higher probability for students who took MI-Phys1 to be connected to each other. Whereas, for the O Group, we see a negative pre-IHI, which indicates heterophily among the students who did not take MI-Phys1. This means that students in the O Group are more likely to be connected to students in the M Group. As we move to the post-IHI in both groups, we see a decrease in the IHI, moving toward zero. Particularly looking at the M Group, we see that the IHI is cut in half. The decrease in IHI from pre- to post- suggests that while there was a clear distinction in representation use between the M and O Groups at the start of the semester, there is far less separation between the groups by the end of the semester.

B. Walk-trap community detection

Using the walk-trap algorithm from Section IIIB, we found that both the pre- and post-networks had two communities. In the pre-network, there were 39 nodes in the first community and 30 nodes in the second, with a modularity of 0.0393. However, in the post-network, there were 42 nodes in the first community and 18 in the second, with a modularity of 0.0190.

The difference in the number of nodes in each group between the pre- and post-network suggests that over the course of the semester, the students are becoming more of a single community and using many of the same representations. We also see this in the drop in modularity from pre to post, which indicates that the two communities are harder to distinguish in the post-network compared to the pre-network.

When we ran the linear model, WalkTrapCommunities ~ Physics1Groups. We found that there was a significant(p << 0.001), positive correlation of 0.485 in the pre-network between the communities detected and the type of Physics 1 course. In contrast, there was no significant correlation between the detected communities and the Physics 1 course in the post network (summarized in Table II).

The linear model results tell us a similar story to those from the homophily analysis. We see that for the pre-network, the course that students took for Physics 1 was a significant factor in determining the community that they belonged to, in how the students were connected to each other, and thus, in what representations they chose to use. The lack of a significant correlation in the post-network indicates that while there were still two communities detected, the student’s Physics 1 course is no longer an influence in their position within the network. This corroborates the finding that by the end of the semester, the students are using similar representations despite different Physics 1 courses.

V. CONCLUSION

From both the homophily analysis and the walk-trap community analysis, we found that the type of Physics 1 course students took significantly impacted the representational choices that they made. When communities were determined in the pre-network, the student’s Physics 1 course was a significant factor, with a correlation of 0.485. Furthermore, there was a higher IHI in
TABLE II. This table shows the results of the linear regression for the walk trap communities as predicted by the Physics 1 course in both the pre and post networks. The $\beta$ value listed shows the correlation with it’s associated $p$-value. The adjusted $R^2$, $F$-statistic, and it’s associated degrees of freedom and $p$-value are also shown.

<table>
<thead>
<tr>
<th>Network</th>
<th>$\beta$</th>
<th>p-value ($\beta$)</th>
<th>Adj.$R^2$</th>
<th>$F$-statistic</th>
<th>Deg. of Freedom.</th>
<th>p-value ($F$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre</td>
<td>0.485</td>
<td>$p &lt; 0.001$</td>
<td>0.227</td>
<td>21.0</td>
<td>(1, 67)</td>
<td>$p &lt; 0.001$</td>
</tr>
<tr>
<td>Post</td>
<td>0.0938</td>
<td>0.438</td>
<td>-0.00665</td>
<td>0.611</td>
<td>(1, 58)</td>
<td>0.438</td>
</tr>
</tbody>
</table>

the pre-network, indicating that students who were in MI-Phys1 tended to be more connected to each other.

However, in the post network, the IHI for both M and O Groups moved toward zero, which shows that by the end of the semester, there was less preference for one group over the other. This suggests that by the end of the semester, there was much less of a distinction between the two groups based on their Physics 1 course. These results were also supported by the drop in network modularity from pre to post, which indicates that the communities were harder to distinguish. In addition, the student’s Physics 1 course was no longer a significant factor in determining the students’ communities.

While students may be entering MI-Phys2 with a variety of experiences from Physics 1, by the end of MI-Phys2, there is much less of a dependence on the Physics 1 course in terms of the students’ representational choices. This means that students are leaving MI-Phys2 with a common set of representations and tools for approaching problems, regardless of how they entered the class. We can conclude that the MI-Phys2 course creates representational continuity among students despite the variety of preparation from Physics 1.

MI courses were designed to provide students with a varied, functional toolbox of representations to use in their future scientific endeavors and to communicate with other scientists. Students who took both MI-Phys1 and MI-Phys2 have spent a full year (as opposed to a single semester) in this education environment where the focus was explicitly on building up this toolbox, establishing a strong foundation in representational competence. Seeing the homogenization of students in MI-Phys2, despite their varied background in Physics 1, shows that students are leaving MI-Phys2 with the representational tools to be successful in science.

The future directions of this work will then focus on what are the specific representations that students from both MI and other Physics 1 backgrounds use in their problem solving. We have seen evidence that they have the same tool box of representations, but we want to know what are the individual tools they are using. In addition, we want to determine how the context of a problem (NM or E&M) influences students’ representation choices.

ACKNOWLEDGEMENTS

We thank the PER group at FIU for their feedback on this project, as well as Dr. Geoff Potvin for allowing us to collect data in his course. This work was supported by NSF Grant # DUE 1140706 and #DGE 1038321.