

Connecting the dots: Student social networks in introductory physics labs

Cole Walsh, Daniyar Kushaliev, and N.G. Holmes

Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY 14853, cju295@cornell.edu

(Dated: July 29, 2020)

Students' positions within the social network of a physics classroom have been shown to correlate with students' sense of belonging, performance, and persistence in physics. Current research in PER aims to understand how different types of active learning classrooms promote the development of students' social networks. In this work, we begin to examine how these networks develop in introductory physics labs where there is typically ample space and freedom for students to interact with their peers and build a community of learners. The primary goal of this paper is to explore the effectiveness of using video to capture information about student networks in this setting. We use two distinct methods of video coding, which we denote as the "scan" and "skip" methods, to capture two distinct types of interactions between students in our labs: interactions between students from different lab groups and interactions between students from the same lab group. We find that both methods are robust and reliable when used by independent coders, supporting the continued use of these methods in future analyses, but that each method has associated advantages and limitations that should be considered along with the research question being addressed. We use these two methods to explore how students' social networks develop over time.

I. INTRODUCTION

Students' learning in the classroom is impacted by the social and emotional climate of the course [1]. Previous work in physics education research (PER) found a positive correlation between students' grades and their social interactions after controlling for prior content knowledge [2], supporting this intrinsic social nature of learning. Social network analysis (SNA) is a rapidly growing field in PER that provides an alternative lens for examining how learning happens in the classroom through a participationist perspective [3, 4] where "learning [...] occurs as people participate in the sociocultural activities of their community, transforming their understanding, roles and responsibilities as they participate" [4].

Researchers have also linked students' positions in social networks to other measures of success, besides learning, that are valuable to PER. Current theories treat interactions between students and other members of an institution, especially in students' first year of college, as a critical element of the student retention model [5]. Sawtelle *et al.* previously found that vicarious learning experiences, where students compare their own achievement with that of a peer, are important for retaining women in first semester introductory physics courses [6], while Dou *et al.* found that student popularity was positively correlated with overall self-efficacy [7]. These studies illustrate the importance of social integration, not only for learning, but also for the development of students' self-efficacy and interest in pursuing physics.

These studies predominantly used SNA with traditional lecture and modeling instruction classrooms [8], and recent work has moved to examining different types of active learning, such as Peer Instruction [9]. Student interactions are also pervasive in lab courses, and so we use SNA here to examine social networks in physics labs. Labs are a common component of most introductory physics courses and provide ample opportunities for students to work with peers and build a community of learners. Previous work examined how physics lab instruction affected students' attitudes towards experimental physics [10, 11] and development of critical thinking skills [12]. SNA provides an alternative lens for interpreting these results through a participationist perspective.

The goal of this work is not to establish how students' social integration in physics labs affects outcomes such as retention, self-efficacy, attitudes, or critical thinking skills; these investigations will be reserved for future work. We instead aim to lay the groundwork for these future studies by investigating how best to collect and analyze social network data for students in physics labs.

Prior work in PER has primarily relied on students' self-reporting interactions to collect SNA data [13]; we investigate here the use of video to collect information about student interactions instead. Recent work used interval-based coding [14, 15] of video to examine student interactions. We found that this method of video coding, which we denote as the "skip" method, was useful

for capturing interactions between members of the same lab group, but had limited use in capturing interactions between students from different groups. To address this limitation, we developed a second video coding method, which we refer to as the "scan" method, which incorporates all time segments in the coding process, rather than skipping chunks of time. We found that this method, conversely, was useful for capturing interactions between students from different groups, but ineffective for evaluating within-group interactions. In the rest of this paper, we describe the "scan" and "skip" methods in greater detail, evaluate the reliability of these methods, and use them to investigate the development of students' social networks over multiple lab sessions.

II. METHODS

A. Data sources

We collected video of students in one section of an introductory mechanics lab at Cornell University. About 95% of students in these labs intended to major in engineering or another science besides physics. The lab curriculum emphasized open-ended investigations designed to engage students in experimentation. We used video of three sequential lab sessions here. Students, in groups of two or three, carried out two separate lab experiments. The first experiment took place over weeks 1 and 2, then students switched groups and worked on a second experiment during week 3. Groups typically worked near their own lab bench, which was physically separated from other groups' lab benches. We used video collected from two cameras positioned in opposite corners of the room to capture the entirety of the lab room and student interactions. We recorded student interactions using two complementary methods that we describe below.

B. Video coding

In our first method ("skip" method), we coded data in a similar fashion to Refs. [14, 15]. We separated lab sessions into two minute intervals, then examined the first 15 seconds of each interval to see which students talked to each other. We used two minute intervals, as in Ref. [14], while we found that 15 seconds was generally sufficient time to observe interactions. We only analyzed instances where *both* students talked to each other within the 15 second window, though one-way communication between students (i.e., without reciprocity from the other student) were recorded as well and may be used in future projects. We recorded interactions in units of counts (i.e., number of times that two students talked to each other).

In our second method ("scan" method), we scanned through videos of each lab session and recorded when students moved to a different lab bench. As we demonstrate below, students spend most of their time working in close proximity to their group's lab bench, but do occasionally move to get equipment from elsewhere in the room, talk to the TA, or visit another group at a different

lab bench. For the purposes of this preliminary work, we were only interested in student-student interactions, so we only recorded instances where a student stopped at a lab bench to interact with other students. We did not track which individual students talked to each other; we considered students that were physically located at the same lab bench to have interacted for the full period of time that the students were at the same location. We recorded interactions in units of time (e.g., minutes).

The scan method is advantageous for capturing all of the inter-group interactions, some of which were missed by aliasing the videos into two minute intervals in the skip method. We cannot determine with the scan method, however, who talks to who when students are at the same lab bench. The skip method can fill in some of this information by capturing which students talk to each other. We view the skip method as being particularly useful for capturing intra-group interactions: how often group members interact with each other and how these interactions vary across groups, sections, and time.

Lab sessions were roughly two hours long, so the maximum amount of interaction between two students for these two methods was roughly two hours or 60 counts, for the scan and skip methods, respectively. It took approximately two hours to code a lab session using the scan method, and approximately six hours using the skip method.

C. Social Network Analysis

We constructed student social networks by treating each individual student as a *node*. We then connected the nodes with weighted *edges* equal to the length of time that students were at the same lab bench (scan method) or the number of times that we coded students speaking to each other (skip method). For the remainder of this paper, we will refer to this representation of student social networks with nodes and edges as a *graph*. We provide example graphs in Fig. 1 for week 2 coded with both methods.

The scan method was used primarily to investigate inter-group interactions, which, as mentioned above, occurred much more infrequently than intra-group interactions. In week 2, for example, no student spent more than 135 seconds at another group’s lab bench. The edges between students from the same group were much stronger, then, and washed out the inter-group interactions in our data. To overcome this background signal, we set the weight of all intra-group edges in the graphs (denoted by dotted lines in Fig. 1a) equal to one second and kept all inter-group edges at the coded weights. We used a weight of one second, rather than removing these edges entirely because we wanted to preserve the total number of edges in the graph, but suppress the signal from intra-group interactions. We used the skip method primarily to investigate intra-group interactions, but we kept all inter-group edges in our analysis so we could compare how both methods measured inter-group interactions.

We first evaluated the interrater reliability (IRR) of our two video coding methods by having two of the authors independently code a full lab session (week 2) using each of the methods described in Sec. II B. There are no well-established methods for measuring IRR with SNA data, so we adapted a graph similarity measure, the *Hamming distance* (H) [16], to compare the graphs obtained by the two independent coders. H measures the number of unweighted edges that would need to be added or removed from one graph to arrive at the second graph. We normalized H by dividing by the total number of possible edges in our graphs (i.e., $n(n-1)/2$, where n is the number of nodes). H is bounded between zero and one, with zero indicating identical graphs and one indicating complementary graphs.

There are two main issues with using H as a measure of IRR. First, H takes into account all possible edges in the graph, including those not coded by either author, so $1 - H$ should be thought of as an upper bound on the IRR. Second, H neglects the weights of the edges. We calculated a second IRR index to overcome these issues, which we defined in two forms, an unweighted and weighted version:

$$R = \frac{\text{number of edges coded by both coders}}{\text{number of edges coded by at least one coder}} \quad (1a)$$

$$R_w = \frac{\text{sum of weights coded by both coders}}{\text{sum of weights coded by at least one coder}} \quad (1b)$$

These metrics only consider interactions coded by at least one author, ignoring edges that neither author coded, and should be viewed as lower bounds on the IRR.

Beyond evaluating and comparing our two methods of video coding, we also examined the development of student social networks over time. We examined graphs using graph-level metrics that characterize the topology (i.e., how the network is arranged and how nodes and edges relate to each other) of the graph as a whole. These metrics can be used to compare two or more graphs in their entirety. We will make use of node-level metrics in future studies to understand how students’ positions in the network relate to other outcome measures.

We used *density* (Δ) to characterize the topology of our graphs here. Δ is the fraction of edges present out of the the total number of possible edges [17]:

$$\Delta = \frac{\text{number of present edges}}{\text{number of possible edges}}. \quad (2)$$

For undirected graphs, such as the ones we use here, the denominator is equal to $n(n-1)/2$, where n is the number of nodes. We also used a modified density measure for weighted graphs that took into account the weights of the edges in the graph [18]:

$$\Delta_w = \frac{\sum \text{weight}(u,v)}{\text{number of possible edges}}, \quad (3)$$

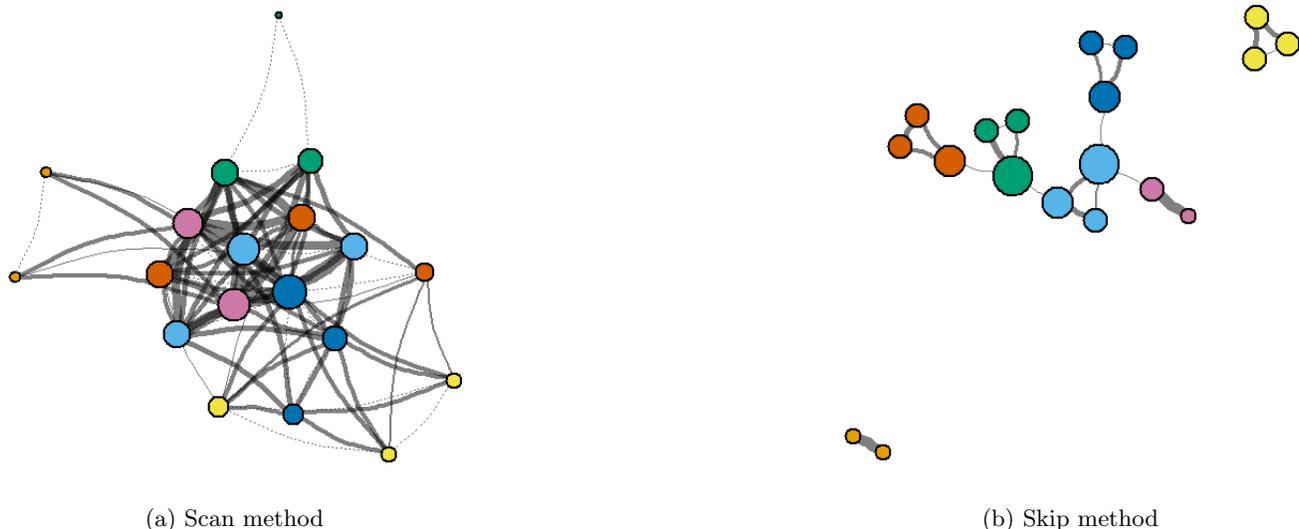


FIG. 1: Graphs for week 2 obtained using both methods of video coding. Node sizes are proportional to the number of students that a student interacted with (i.e., the number of edges connected to the node) and node colors indicate students that worked in the same group for the session. Edge widths are proportional to the length of time two students spent at the same lab bench (scan method) or the number of interactions between two students (skip method). In Fig. 1a, we have set the weight of all intra-group edges (represented by dotted lines) equal to one second, so that the inter-group edges are more dominant in the analysis.

where we sum over the weights of all pairs of nodes, u and v , in the numerator.

Δ has the same interpretation for graphs obtained from both methods: roughly the average fraction of students that each student interacted with in the lab. We used Δ to compare the *number* of interactions captured by the two methods. Δ_w , on the other hand, has different interpretations depending on which video coding method produced the graph. For the scan method, Δ_w is roughly the average *amount of time* that a student interacted with each other student, while for the skip method, Δ_w is approximately the average *number of times* that a student interacted with each other student. We tracked these density values over time to evaluate the evolution of student connectivity in labs and to get a sense of the variability in graph topology.

III. RESULTS AND DISCUSSION

A. Reliability of video coding

The Hamming distance and reliability indices between graphs of week 2 obtained by two independent coders for each method are presented in Table I. The Hamming distance is 0.035 and 0.018 for the scan and skip methods, respectively, indicating that of all possible edges in the graph, the two coders disagreed on the existence of 3.5% of edges when using the scan method and 1.8% of edges when using the skip method. There were 171 possible edges in this graph, so this disagreement corresponds to six and three edges, respectively.

As mentioned in Sec. II C, $1 - H$ (0.965 and 0.982

TABLE I: Hamming distance and reliability indices between graphs of week 2 obtained by two independent coders using both video coding methods.

	Scan method	Skip method
Hamming distance	0.035	0.018
Reliability (unweighted)	0.931	0.857
Reliability (weighted)	0.854	0.829

for the scan and skip methods, respectively) can be thought of as an upper bound on the IRR. We compute a lower bound by calculating the reliability index defined in Eq. 1a. These values, 0.931 and 0.857 for the scan and skip methods, respectively, also indicate that there was strong agreement between the two coders as to *which* edges exist in the graph. Finally, our weighted reliability index (see Eq. 1b) is equal to 0.854 and 0.829 for the scan and skip methods, indicating that more than 80% of the interaction times or counts coded by at least one author were coded by both authors.

We do not have benchmarks for what constitutes “good” reliability as this is the first time the Hamming distance has been used in this way, to our knowledge. Other measures of IRR with range between zero and one, and used with other forms of data, such as Cohen’s κ , typically treat values greater than 0.8 as extremely strong agreement [19]. We argue, then, that the graphs obtained here by independent coders agree very well, indicating that both methods of video coding are robust and reliable.

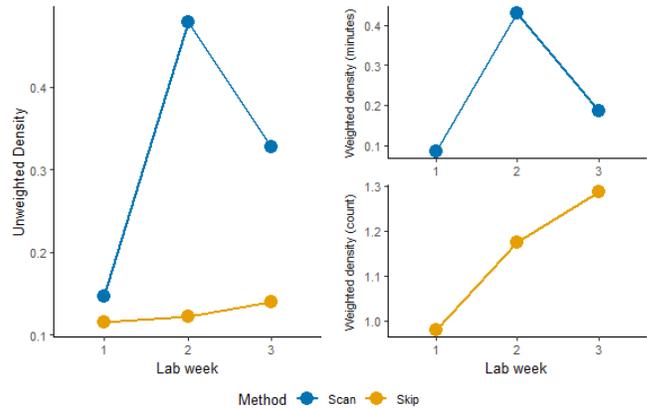
B. Graph density

Figure 1 illustrates that the two video coding methods capture different information about students' interactions. The graph produced using the skip method (see Fig. 1b) contains nodes that are clustered in students' lab groups and reveals differences in interactions among students from the same group. There are only four inter-group edges in Fig. 1b, however, compared to 65 in Fig. 1a. Some of these additional 61 edges arose because of the less restrictive definition of an interaction used in the scan method, where students at the same lab bench were coded as having interacted. Other inter-group interactions were missed using the skip method due to the short nature of these interactions and the aliasing into two minute intervals. These inter-group interactions were captured by the scan method, however, and provide important information about the larger social structure of the lab. Two groups are isolated in Fig. 1b, for example, but are connected to other groups in Fig. 1a, indicating that short inter-group interactions involving these groups were missed using the skip method.

We examined the evolution of Δ (Eq. 2) over three lab sessions in Fig. 2. We found that Δ for graphs obtained using the scan method were generally larger than those obtained using the skip method because of the reasons mentioned above: the more restrictive definition of an interaction and aliasing into two minute intervals in the skip method. We also found, focusing on inter-group interactions and the scan method, that Δ increased from week 1 to week 2 with a subsequent dip with the beginning of a new experiment in week 3. Week 2 generally involved more open-ended investigations than weeks 1 and 3, so students may have been more compelled to compare with or seek direction from other groups. We intend to investigate whether this trend holds later in the semester and for other sections.

We cannot compare Δ_w (Eq. 3) for graphs obtained by the two coding methods because the edge weights used have different units (i.e., minutes versus count). We can, however, compare the *evolution* of Δ_w . As with Δ , we found that Δ_w increased from week 1 to week 2 (in conjunction with both coding methods), indicating that students spent more time talking to their group members, on average, and more time at other groups' lab benches, on average, in the second week of the lab. We found that Δ_w decreased between weeks 2 and 3 for graphs obtained using the scan method, but increased for graphs obtained using the skip method. These results indicate that the average number of intra-group interactions increased despite the length (and number) of inter-group interactions decreasing. This result could indicate that students are becoming more active participants in the lab, but data from later lab sections would be required to determine whether students are only becoming more active within groups at the expense of inter-group interactions. We, again, intend to investigate whether this holds later in the semester, and also whether Δ_w evolves similarly over

FIG. 2: Evolution of graph density over the first three lab sessions. Weighted densities are measured in different units for the two methods and should not be directly compared.



time using both methods.

IV. LIMITATIONS AND FUTURE WORK

In this paper, we compared the effectiveness of two methods of video coding for capturing student interactions in an introductory physics lab. Our results indicate that both methods are highly reliable and provide complementary information. There was a difference in the efficiency of these methods, however, which has implications for the research questions that can be answered at larger scales. With future work, we intend to examine different labs at multiple institutions over time, so the limited efficiency of the skip method will inhibit our ability to scale up this analysis to answer research questions specific to intra-group interactions. The scan method is effective for capturing inter-group interactions, on the other hand, but does not capture information about who talks to who. A hybrid of both methods that codes who talks to who between groups could efficiently capture the inter-group interactions at a finer scale.

In our preliminary analysis, we found that the density of students' social networks generally increased over the short time period we examined. In follow-up studies we will examine other graph-level features such as how well students cluster together and how balanced interactions are within groups to understand how graph topologies are different or similar across different institutions and lab pedagogies. Building on prior work, we will also investigate how students' positions in the lab social network correlate with other outcome measures such as attitudes about experimental physics and critical thinking skills.

ACKNOWLEDGMENTS

We wish to thank the Cornell Discipline-based Education Research group for helpful feedback on this manuscript.

-
- [1] Susan A Ambrose, Michael W Bridges, Michele DiPietro, Marsha C Lovett, and Marie K Norman, *How learning works: Seven research-based principles for smart teaching* (John Wiley & Sons, 2010).
- [2] Jesper Bruun and Eric Brewes, “Talking and learning physics: Predicting future grades from network measures and force concept inventory pretest scores,” *Physical Review Special Topics-Physics Education Research* **9**, 020109 (2013).
- [3] Eric Brewes, Laird Kramer, and Vashti Sawtelle, “Investigating student communities with network analysis of interactions in a physics learning center,” *Physical Review Special Topics-Physics Education Research* **8**, 010101 (2012).
- [4] Barbara Rogoff, Eugene Matusov, and Cynthia White, “Models of teaching and learning: Participation in a community of learners,” *The handbook of education and human development*, 388–414 (1996).
- [5] Vincent Tinto, “Research and practice of student retention: What next?” *Journal of College Student Retention: Research, Theory & Practice* **8**, 1–19 (2006).
- [6] Vashti Sawtelle, Eric Brewes, and Laird H Kramer, “Exploring the relationship between self-efficacy and retention in introductory physics,” *Journal of research in science teaching* **49**, 1096–1121 (2012).
- [7] Remy Dou, Eric Brewes, Justyna P Zwolak, Geoff Potvin, Eric A Williams, and Laird H Kramer, “Beyond performance metrics: Examining a decrease in students’ physics self-efficacy through a social networks lens,” *Physical Review Physics Education Research* **12**, 020124 (2016).
- [8] Eric Brewes, Laird H Kramer, and George E O’Brien, “Changing participation through formation of student learning communities,” in *AIP Conference Proceedings*, Vol. 1289 (American Institute of Physics, 2010) pp. 85–88.
- [9] Kelley Commeford, Eric Brewes, and Adrienne Traxler, “Characterizing active learning environments in physics: network analysis of peer instruction classroom using ergms,” in *Physics Education Research Conference 2019*, PER Conference (Provo, UT, 2019).
- [10] Bethany R Wilcox and HJ Lewandowski, “Developing skills versus reinforcing concepts in physics labs: Insight from a survey of students’ beliefs about experimental physics,” *Physical Review Physics Education Research* **13**, 010108 (2017).
- [11] Bethany R Wilcox and HJ Lewandowski, “Open-ended versus guided laboratory activities: Impact on students’ beliefs about experimental physics,” *Physical Review Physics Education Research* **12**, 020132 (2016).
- [12] Cole Walsh, Katherine N Quinn, C Wieman, and NG Holmes, “Quantifying critical thinking: Development and validation of the physics lab inventory of critical thinking,” *Physical Review Physics Education Research* **15**, 010135 (2019).
- [13] Eric A Williams, Justyna P Zwolak, Remy Dou, and Eric Brewes, “Linking engagement and performance: The social network analysis perspective,” *Physical Review Physics Education Research* **15**, 020150 (2019).
- [14] Katarzyna E. Pomian, Justyna Zwolak, Eleanor Sayre, Scott Franklin, and Mary Bridget Kustus, “Using social network analysis on classroom video data,” in *Physics Education Research Conference 2017*, PER Conference (Cincinnati, OH, 2017) pp. 316–319.
- [15] Christopher A. Hass, Florian Genz, Mary Bridget Kustus, Pierre-P. A. Ouimet, Katarzyna E. Pomian, Eleanor Sayre, and Justyna Zwolak, “Studying community development: A network analytical approach,” in *Physics Education Research Conference 2018*, PER Conference (Washington, DC, 2018).
- [16] Giuseppe Jurman, Roberto Visintainer, Michele Filosi, Samantha Riccadonna, and Cesare Furlanello, “The hinged metric and kernel for network comparison and classification,” in *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (IEEE, 2015) pp. 1–10.
- [17] Remy Dou and Justyna P Zwolak, “Practitioner’s guide to social network analysis: Examining physics anxiety in an active-learning setting,” *Physical Review Physics Education Research* **15**, 020105 (2019).
- [18] Steve Horvath, *Weighted network analysis: applications in genomics and systems biology* (Springer Science & Business Media, 2011).
- [19] J Richard Landis and Gary G Koch, “The measurement of observer agreement for categorical data,” *biometrics*, 159–174 (1977).