

Examining the effects of lab instruction and gender composition on intergroup interaction networks in introductory physics labs

Meagan Sundstrom¹, David G. Wu¹, Cole Walsh¹, Ashley B. Heim², and N. G. Holmes¹

¹Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA

²Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853, USA



(Received 7 September 2021; accepted 1 December 2021; published 4 January 2022)

Understanding social interactions among students comprises a rich area of physics education research. Here we focus on the social interactions in introductory physics laboratories (labs). Most existing research in such contexts focuses on within-group social dynamics, however, we argue that interactions between different lab groups are just as valuable, especially in nontraditional (reformed) labs where students have more control over the experimental designs. Using video recordings of lab sessions and social network analysis methods, we examined whether and for how long groups interact with each other, comparing across three offerings of an introductory mechanics course. We observed significant variability in the interactions across lab sessions and no clear pattern in how strongly lab groups were connected in the networks between the different courses. More prolonged intergroup interactions, however, occurred in three of the four reformed lab sections, as compared to a traditional lab section and a reformed lab section that took place in the evening. We also developed a group-level social role classification scheme based on groups' interaction patterns, identifying four roles: noninteractors, information seekers, responders, and mutual interactors. We found that groups in the traditional lab section disproportionately acted as noninteractors, indicating that many groups in this lab did not engage in intergroup interactions at all. In contrast, groups in reformed lab sections took on the remaining three roles more frequently. We also found possible relationships between the distributions of the social roles within each type of lab by the groups' gender compositions. All-male groups took on interactive roles more often than all-female and mixed-gender groups in the reformed lab sections, but not in the traditional lab. Results indicate that the amount of time students spend on intergroup interactions may depend on the lab curriculum itself, the time of day of the section, and/or the student population of the section. Still, instructors can encourage collaboration between lab groups through direct prompts, but such prompts need to be repeated throughout the semester.

DOI: [10.1103/PhysRevPhysEducRes.18.010102](https://doi.org/10.1103/PhysRevPhysEducRes.18.010102)

I. INTRODUCTION

Much of educational theory is built on the premise that learning is social and occurs through a participationist perspective [1–3]. Constructivist theory similarly posits that learners construct knowledge in the context of their specific social surroundings [4,5]. In tandem, education researchers in physics and other science disciplines have demonstrated that interactions with peers enhance undergraduate students' academic achievement [6–9], self-efficacy [10], and persistence [11–13], and also decrease undergraduate students' physics anxiety [14].

With clear benefits from student interactions, a plethora of evidence-based instructional styles that call for students to engage in such interactions have emerged in the

education research literature. These developments warrant more detailed analyses of how students interact with peers in specific learning environments. Several studies have compared the interaction networks of physics students in different active learning environments [15–17]. For instance, one study found that students in a Modeling Instruction physics course (where students work in small groups to construct models, beginning with qualitative descriptions of phenomena and building up to quantitative representations [18]) formed an interactive community, while students in a traditional lecture course (with no structured group work) were more isolated, with little to no social connections between peers [15]. Another study compared student interactions across various active learning formats in physics using positional analysis [17]. They found that students who start the semester positioned in groups of close peers form stable interaction networks throughout the duration of the course. This finding held true across multiple instructional styles.

Evidently, it is important to examine how students interact with peers during different kinds of instruction.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Here we focus on student interactions during instructional laboratories (labs). Because of the nature of experimental tasks and small-enrollment sections, students engage heavily in group work in introductory physics labs. Students place high value on socialization and help-seeking opportunities in such labs as well [19,20].

However, existing research on social interactions in introductory physics labs is largely focused on within-group gender dynamics [21–27]. While these studies are important for understanding the social dynamics that occur among students working in the *same* lab group, students may also benefit from observing and engaging in discussions with peers in *other* lab groups. For example, collaborating with other groups allows for the exchange of information (e.g., experimental procedures or results), clarification of a point of misunderstanding, or fulfilling a lack of equipment. In line with the constructivist point of view, students build their understanding and inform their experimental decisions through interactions with others both within and beyond their own group (i.e., their complete social surroundings). As Lowe *et al.* describe, “being able to eavesdrop on related conversations, notice the issues confronting other students, and overhear the questions they are asking to the instructor, can all play a role in assisting the learning process” [28] (p. 53). In other words, students’ engagement in and understanding of lab activities is improved when they can hear about alternative approaches to or perspectives on the ongoing investigation.

Broadening the current understanding of within-group interactions in the physics education research (PER) literature to the interactions between different lab groups is thus necessary to further grasp the role of socialization in labs. The aim of the current study is to examine the extent and patterns of intergroup interactions in different kinds of introductory physics labs. We apply methods of social network analysis (SNA) to video data of five separate lab sections in three offerings of an introductory physics lab course.

A. Interplay of lab curriculum and social interactions

Research has shown that the instructional style of a lecture course impacts the amount and structure of interactions among students [15–17]. The same is likely true of labs. That is, the nature of lab tasks likely affects the motivation or number of opportunities students have to engage in interactions with other groups. For instance, groups in a traditional physics lab follow the same experimental procedure, leaving little advantage to seeking information about others’ approaches.

Physics education researchers, however, have prompted a move away from traditional, model-verification labs and toward inquiry-based labs that focus on experimentation skills rather than content reinforcement [29–34]. While traditional labs ask students to solve closed and structured problems, reformed labs promote more complex conceptions about the nature of science by having students make

and reflect on their own decisions [31]. These nontraditional labs better support students’ agency and engagement in scientific abilities [33–39] and views about experimental physics [33,40,41] compared to traditional labs. Such reformed labs also place higher emphasis on teamwork, collaboration, and communication skills [30,42]. Reformed labs may require students to select their own research questions and design their own procedures, often resulting in different lab groups approaching the lab activities in different ways or arriving at different findings [43]. While students in reformed labs may get a lot of help from their immediate group mates, they are likely going to succeed more by exploring the diverse methods and findings of other groups.

One goal of the current study is to directly compare the impact of traditional and reformed physics lab curricula on intergroup interactions. Researchers in biology education have conducted a few studies in this vein, examining interactions in course-based undergraduate research experiences (CUREs). Similar to reformed physics labs, students in biology CUREs engage in iterative experimentation to address research problems with no predetermined or known answers [44]. One study found that students in CUREs discuss their labwork with friends, professors, family members, and other students more than students in traditional biology labs [45]. Another study used recorded video data and SNA to examine interactions in different biology lab courses [46]. They found that intergroup interactions are more frequent in CUREs than in traditional biology labs. We predicted similar results for our study: given that the nature of reformed physics labs is similar to that of CUREs, intergroup interactions may be more prevalent in reformed physics labs than in traditional physics labs.

B. Gendered social dynamics and peer perceptions in physics

Understanding social interactions in physics labs undoubtedly requires attention to both gender dynamics and gendered peer perceptions. First, prior research demonstrates that within-group peer interactions in physics labs are subject to inequities. Such studies unanimously found gendered task division within mixed-gender lab groups: women typically write up the lab report or supervise the group, while men typically handle the equipment or analyze data [21–27]. As a specific example, Doucette *et al.* [24] conducted interviews and ethnographic observations to inform a new framework for equity in introductory physics labs. They defined four archetypes or roles that students may take on in this setting, which depend on a student’s gender and level of initiative. They proposed that men with more initiative are “tinkerers” who handle the experimental equipment and women with more initiative are “Hermione” archetypes that take a supervising role. On the contrary, a male student with low initiative may be considered a “slacker,” while a female student with low

initiative might be considered a “secretary.” This work provides an evidentiary basis for the different roles students take on *within* their lab groups. Similarly, we might expect different types of roles to emerge at the group level when considering intergroup interactions.

One goal of the current study, therefore, is to characterize group-level roles and determine whether these roles are related to a group’s gender composition. Interactions between groups are likely sensitive to peer perceptions, given the agency students have in whether or not to engage in such interactions. For example, individuals tend to work with others they view as skilled or knowledgeable [47]. However, a gender bias within peer perceptions has been observed across science, technology, engineering, and math (STEM) disciplines [48,49]. These studies found that students view male peers as stronger in the course material than female peers, even when women outperform men in the course. Perceiving men as more knowledgeable in the course than women might lead students to interact with all-male or mixed-gender groups more than all-female groups.

Another study evaluating peer perceptions [50], however, found contrasting results with respect to the prevalence of gender bias. They attributed the different outcomes to the differences in the students’ course level and majors (an upper-level, small-enrollment mechanical engineering course versus an introductory, large-enrollment biology course) and demographics (a majority-male course versus a gender-balanced course) of each study’s sample. A similar phenomenon may also arise when examining the patterns of student interactions in labs serving different populations. Our study, therefore, aims to determine whether different gendered interaction patterns emerge in the different courses.

C. Current study

As physics instructors shift toward implementing lab curricula that develop students’ experimentation skills, it is critical that we understand the impacts of these labs on student collaboration. In this study, we examine the intergroup networks of introductory physics labs using SNA to quantify interactions between groups. Our two research questions are as follows:

1. How do the structural characteristics of intergroup networks in three different introductory physics lab curricula compare and evolve over a semester?

2. What are the social roles that groups in introductory physics labs take on and are the distributions of these roles different across groups of varying gender composition and across groups in three different lab curricula?

To address these questions, we analyzed video of five introductory physics lab sections, spanning three different lab curricula, for the entirety of a semester.

In the remainder of the paper, we present our study as follows: In the next section, we describe the characteristics of the analyzed lab sections, our video coding procedures, relevant aspects of the SNA methodology, and our methods of analyzing the data. In Sec. III, we present our findings related to group-level networks (research question 1) and social roles (research question 2). We end with a synthesizing discussion of our results in Sec. IV.

II. METHODS

In this section, we first characterize the lab sections included in the study and describe our methods of collecting and coding video of the labs to produce network data. Then, we present relevant principles and definitions related to SNA. We finally describe how we analyzed the data to address each research question.

A. Data collection

Our data come from five introductory physics lab sections at Cornell University. Table I summarizes the five sections by course, lab curriculum, and number of enrolled students. Table II summarizes the self-reported demographics of students in each section. In all cases, the sections were chosen because all students and the instructor consented to being video recorded for research.

Sections 1, 2, and 3 were part of an introductory, calculus-based mechanics course intended for physics majors. Sections 4 and 5 were part of an introductory, calculus-based mechanics course intended for engineers and other STEM majors. As summarized in Table II, most students in each section identified as White and/or Asian or Asian American. Sections 1 through 4 had a majority of men, while section 5 had a majority of women. In all sections, most students were first-year students, though a significant proportion in sections 4 and 5 were second-year students.

TABLE I. Summary of the five lab sections we analyzed, including the physics course, lab curriculum, number of enrolled students, and number of coded sessions.

Section	Course (Major of intended audience)	Lab curriculum	Enrolled students	Coded sessions
1	Intro. Mechanics (Physics)	Early, Traditional	24	7
2	Intro. Mechanics (Physics)	Early, Reformed	23	8
3	Intro. Mechanics (Physics)	Early, Reformed	25	8
4	Intro. Mechanics (Engineering or other STEM)	Recent, Reformed	19	8
5	Intro. Mechanics (Engineering or other STEM)	Recent, Reformed	14	7

TABLE II. Summary of the self-reported race or ethnicity, gender, intended major, and academic year of students in each lab section. Numbers in parentheses are the N values corresponding to the percentages. For gender, students were given the options of male, female, nonbinary, and self-disclose (with a place to type in their response). All students in this study selected either male or female. We note that the computer science major is partly housed in the engineering school at this institution.

		Section 1	Section 2	Section 3	Section 4	Section 5
Race or ethnicity	Asian/Asian American	54% (13)	22% (5)	28% (7)	32% (6)	29% (4)
	Black/African American	4% (1)	0	0	5% (1)	0
	Hispanic/Latino	0	4% (1)	12% (3)	16% (3)	7% (1)
	White	33% (8)	70% (16)	52% (13)	26% (5)	57% (8)
	Multiple races or ethnicities	8% (2)	0	8% (2)	16% (3)	7% (1)
	Other/Unknown	0	4% (1)	0	5% (1)	0
Gender	Male	71% (17)	74% (17)	84% (21)	58% (11)	36% (5)
	Female	29% (7)	26% (6)	16% (4)	42% (8)	64% (9)
Major	Physics/Engineering Physics	58% (14)	39% (9)	36% (9)	0	0
	Other Engineering/Computer Science	13% (3)	30% (7)	24% (6)	69% (13)	57% (8)
	Other STEM	4% (1)	18% (4)	20% (5)	26% (5)	36% (5)
	Unknown	25% (6)	13% (3)	20% (5)	5% (1)	7% (1)
Year	First-year	100% (24)	96% (22)	96% (24)	53% (10)	71% (10)
	Second-year	0	0	4% (1)	42% (8)	29% (4)
	Third-year	0	4% (1)	0	5% (1)	0

Section 1 was a traditional lab, while sections 2 and 3 were reformed labs. Sections 4 and 5 were also reformed labs, however they took place two years later than sections 1, 2, and 3. To distinguish the timing of the courses, we will refer to the lab curriculum of section 1 as “early, traditional,” that of sections 2 and 3 as “early, reformed,” and that of sections 4 and 5 as “recent, reformed.”

In all five sections, students worked together in groups of two to four to complete activities pertaining to a certain topic. Table III summarizes the order of topics covered in each session. All four reformed lab sections employed similar pedagogical activities as described in Refs. [31–34,51]. For example, in Pendulum I and II, students are provided with the canonical model for the period of a pendulum (see Ref. [52]). They are then prompted to design, implement, and refine an experiment to test this model. Students must also decide how to reduce uncertainty in their measurements and how to extend their investigation. The instructor periodically checks in with

each group, providing feedback and suggestions. In subsequent lab sessions, the structure provided by both the activity instructions and the instructor are gradually relaxed. The early, reformed labs differ from the recent, reformed labs through improved teaching assistant (TA) training and general refinements based on observations from implementing the earlier curriculum with many students. The early, traditional labs provided students with detailed instructions to carry out a particular experiment to demonstrate a predefined outcome to support student understanding of the underlying phenomenon (see Refs. [35,53] for more detailed characterizations of traditional lab pedagogy).

Here we focused on the eight sessions that were not the introduction or final presentation sessions because there are no lab tasks or group work in these sessions. Though many previous studies using SNA in education research surveyed students to probe the occurrence of interactions [6–9], surveys can miss a lot of information due to recall bias,

TABLE III. Topics of lab activities for each section analyzed in this study. Sections 2 through 5 (all reformed labs) covered similar topics during the same sessions of the semester, while section 1 (the traditional lab) completed distinct activities.

Session	Section 1 (Early, Traditional)	Sections 2 and 3 (Early, Reformed)	Sections 4 and 5 (Recent, Reformed)
1	1D Motion	Pendulum I	Pendulum I
2	Newton’s laws	Pendulum II	Pendulum II
3	Force laws	Terminal velocity I	Objects in flight I
4	Energy exchanges	Terminal velocity II	Objects in flight II
5	Collisions I	Hooke’s law I	Hooke’s law I
6	Collisions II	Hooke’s law II	Hooke’s law II
7	Angular momentum	Project I	Project I
8	Oscillations	Project II	Project II

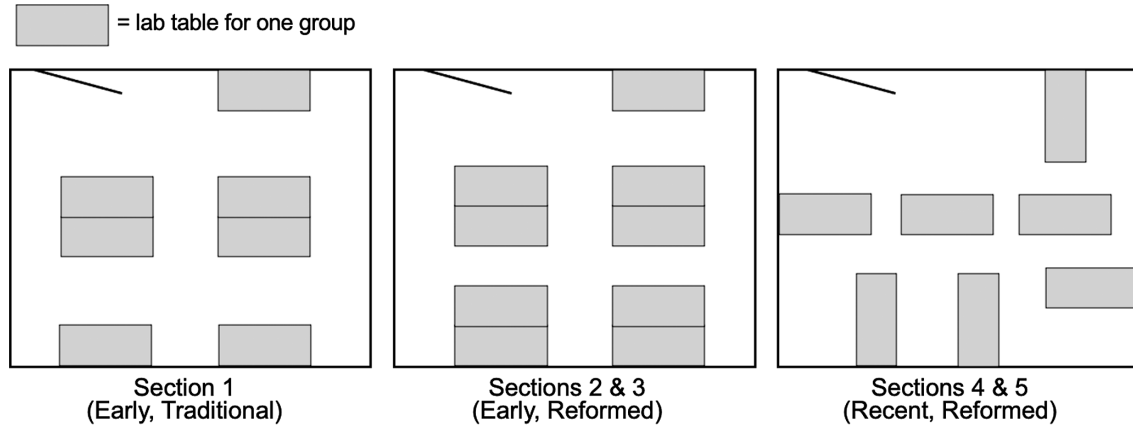


FIG. 1. Schematic of the classroom layouts for the five lab sections analyzed in this study. Solid black lines inside each layout represent the location of the classroom's door. Classrooms for sections 1, 2, and 3 may have more than one lab group at adjacent tables. All lab tables (and lab groups) are physically separated in the classroom for sections 4 and 5.

where students fail to report significant interactions. Therefore, we chose to record all sessions for each section throughout the semester using video cameras overlooking the whole classroom, allowing us to capture all intergroup interactions taking place. Analyses in Refs. [22,33,54–56] use a subset of this recorded video data.

Sessions were not analyzed if we did not have a video recording of that session due to technical reasons or the video recording did not visually capture all of the students in the classroom (i.e., such that we could not confidently infer all of the interactions taking place). Each lab section had at least seven analyzed sessions, as summarized in the last column of Table I. With each session lasting about 2 h, this amounted to a total of roughly 76 h of analyzed video.

The five lab sections were held in three different classrooms with different table layouts, shown in Fig. 1, and were instructed by four different graduate and faculty TAs. The graduate TAs were all experienced TAs, one of whom was familiar with physics education research. The faculty TA was an experienced instructor in the physics department. We discuss the roles of and limitations presented by these additional variables throughout the paper.

B. Video coding

For each session we could analyze, we used video coding methods similar to those described in Ref. [56] to identify interactions between lab groups. This method codes for co-location, assuming that (i) students who are at their own lab table with their own lab group may or may not be interacting with one another at a given time, but (ii) students who physically relocate to another lab group's table temporarily are interacting with that other group. We coded video in sections 4 and 5 exclusively for co-location, i.e., students moving around the classroom to other lab tables, because each lab group sat at a table that was physically separated from all other tables (Fig. 1). In sections 1 through 3, multiple lab groups could be seated at adjacent

tables (Fig. 1) and interact without moving from their own table. For these sections, we coded for both co-location and cross-table interactions between lab groups sitting at adjacent lab tables. When lab groups sat at adjacent tables, at least one group was always facing the camera such that we could use facial movements and physical gestures to infer when the groups were communicating.

For each lab session recording, we used BORIS [57] to scan through the video, focusing on one group at a time. We recorded which groups were interacting, which group initiated the interaction (taken as the group who left their own table for co-location and the group who started talking first for cross-table interactions), and the length of the interaction in units of time. This process was repeated for all lab groups in the session. The coded data for each session then contained the duration and direction of all observed intergroup interactions. This video coding method does not identify within-group interactions.

Four of us contributed to the video coding. We calculated three reliability indices among coders per the methods described in Ref. [56]: the Hamming distance, edge agreement, and weight agreement. The Hamming distance is the number of edges (interactions between groups) for which a pair of coders disagree on their existence as a percentage of all possible edges in the network. The Hamming distance takes into account all possible edges in a network, including those not coded by either individual coder, so this measure subtracted from 1 serves as an upper bound of intercoder reliability. Edge and weight agreement serve as lower bound estimates of reliability. Edge agreement is given by the number of edges coded by a pair of coders as a fraction of the number of edges coded by at least one of the two coders. Weight agreement is given by the sum of edge weights (time duration of interactions) coded by a pair of coders as a fraction of the sum of edge weights coded by at least one of the two coders. The latter two measures can be highly volatile, particularly for sparse networks.

TABLE IV. Ranges (averages) of pairwise intercoder reliability metrics for two lab sessions.

	Dense network	Sparse network
Hamming distance	0.04–0.09 (0.06)	0.03–0.04 (0.04)
Edge agreement	84%–93% (88%)	66%–88% (76%)
Weight agreement	80%–88% (84%)	50%–85% (67%)

For example, if one coder records edges A and B and a second coder records edge A only, then the edge agreement drops to 50% (agreement on one out of two unique coded edges).

Table IV summarizes these three reliability measures for two different lab sessions. As part of previous work [56], three people coded session 2 of section 4, which contained many intergroup interactions (dense network). To be conservative, four co-authors also separately coded session 1 of section 2, which was much less interactive (sparse network). Based on the values shown in Table IV, our measures of sparse networks are less reliable than our measures of dense networks.

C. Methodology: Social network analysis

We drew on methods of SNA to analyze our coded data. SNA is situated in the participationist view of learning [3] and provides a useful methodology for illustrating and quantifying social interactions during instruction. Here, we only provide details about aspects of SNA relevant to the current study. Table V summarizes important SNA terms and definitions for easy reference. For those interested, Refs. [14,58,59] offer holistic, approachable overviews of SNA and means of applying the methodology in education research.

Networks are composed of two main features: *nodes* and *edges*. Nodes represent actors, in our case student lab groups. Edges form the structure of connections among a set of nodes, in our case the interactions between lab groups. Edges can be either *directed* or *undirected*. Edges

of directed networks contain information about the direction of a connection, while those of undirected networks do not. All networks in the current study are directed, indicating which group initiated the interaction.

Once we have formed a network, we can quantify its characteristics either at the individual (node) level or at the network level. At the node level, we can quantify a node's *degree* as the number of edges connected to the node. In our study, a node's degree indicates the number of other groups that node (group) interacted with. For a directed network, we can specify a node's *indegree* as the number of incoming edges connected to the node (number of other groups who initiated an interaction with that node) and the node's *outdegree* as the number of outgoing edges connected to the node (number of other groups with whom the node initiated an interaction). These measures provide information about how central a node is to a network, with a higher degree (more connections) indicating higher *centrality*.

Edges may also be *weighted*, with the weight indicating some measure of intensity of the connection. In our study, we weighted edges by the duration of the interactions. We can quantify a node's *strength* as the sum of the weights of all edges connected to the node. For our directed networks, we can specify the *instrength* of a node as the sum of weights of all incoming edges connected to the node (total time spent on interactions initiated by other groups) and the *outstrength* of a node as the sum of weights of all outgoing edges connected to the node (total time spent on interactions initiated by the group). Similar to degree, a higher strength (stronger connections) indicates higher centrality.

At the network level, we can express the *density* of connections as the fraction of possible edges that we actually observe [14]. Mathematically, the density of a network is given by

$$\text{density} = \frac{\text{number of observed edges}}{\text{number of possible edges}}.$$

TABLE V. Glossary of network terms relevant to this paper.

Term	Definition
Nodes	Actors or entities (in this study, lab groups) connected in some way.
Edges	The structure of connections (in this study, interactions) between nodes. Edges may be weighted, indicating some measure of intensity of connections.
Degree	The number of edges connected to a node. For a directed network, <i>indegree</i> is the number of incoming edges connected to the node and <i>outdegree</i> is the number of outgoing edges connected to the node.
Strength	The sum of weights (in this study, the time durations of interactions) of all edges connected to a node. For a directed network, <i>instrength</i> is the sum of weights of all incoming edges connected to the node and <i>outstrength</i> is the sum of weights of all outgoing edges connected to the node.
Density	The ratio of the number of observed edges in a network to the number of possible edges that could exist with that particular number of nodes. For a directed network of N nodes, $\text{density} = (\text{number of observed edges})/[N(N-1)]$.
Total strength	The sum of all weights of all edges in the network.

For a directed network of N nodes, the number of possible edges (i.e., if every possible pair of nodes were connected by edges in both directions) is $N(N - 1)$. Density values range from 0 to 1 and give an overall sense of the nodes' connectedness, with higher densities corresponding to more strongly connected networks. Finally, we can also determine the *total strength* of the network as the sum of all weights of all edges in the network. The total strength, in our case, indicates the total duration of intergroup interactions in the session, which we can normalize by the overall duration of the lab session and number of lab groups present.

D. Analysis: Intergroup networks

To address our first research question, we used the coded video data to create network diagrams illustrating the occurrence and duration of observed interactions between groups. As a first step, we qualitatively examined the structural characteristics of the networks within and across lab sections to identify general patterns of intergroup interactions in these contexts.

To make comparisons quantitatively, we first calculated and plotted the *unweighted network density* for all coded sessions. The unweighted network density is the number of pairs of groups that we observed as interacting scaled by the total number of possible pairs of groups that could have interacted. We emphasize that the measure is unweighted; it does not take interaction duration into account. The unweighted network density quantifies the connectedness of lab groups, i.e., the diversity of information sharing.

We also calculated and plotted the *normalized total strength* of each lab session, given by the sum of the durations of all observed intergroup interactions divided by the number of lab groups present and the duration of the lab session. This index represents the average fraction of the lab session that each group spent interacting with other groups. Dividing by the number of groups takes into account any differences between sections of varying size. For example, consider two lab sections, one with five groups and one with seven groups, with the same total duration of intergroup interactions. The normalized total strength will be greater for the section with five groups representing that, on average, each group in this section spent more time interacting than the groups in the larger section. In addition, normalizing by the lab duration ensures comparability between lab sessions, as shorter lab sessions provide less time for interactions and longer lab sessions provide more time for interactions. All reformed labs were on average 2 h long, so we used this as the lab length for all sessions of sections 2 through 5. The traditional lab had more variable lengths (all shorter than two hours). For this section (section 1), we used the time from when the TA started the lab until the time the last lab group left the classroom or the camera recording ended (whichever occurred first) as the length of each session. The

normalized total strength provides information about the durations of the observed interactions that are not captured by the unweighted network density. Taken together, the two measures provide a holistic picture of the number and duration of interactions taking place in each lab section, which we then examined over time and across different lab curricula.

E. Analysis: Group-level social roles

To address our second research question, we identified group-level social roles by classifying all $N = 265$ lab groups analyzed as part of this study (where each group in each lab session served as a unique group observation) by numbers and types of intergroup interactions. We used two measures of centrality in our classification: *normalized instrength* and *normalized outstrength*. A group's normalized instrength is the sum of the durations of interactions to which the group responded (another group initiating) divided by the length of the given lab session. High normalized instrength, therefore, means a group spent a considerable fraction of the lab session interacting with other lab groups who approached them. A group's normalized outstrength is the sum of the durations of interactions that the group initiated divided by the length of the given lab session. This measure can be taken as a quantification of a lab group's information seeking behavior, as it only takes into account the interactions in which the group intentionally approached another group. In this way, our social role identification and classification process was grounded in our quantitative video coding data.

We ran a k -means cluster analysis [60] on the two dimensions of normalized instrength and normalized outstrength to get a sense of the groupings. The cluster analysis suggested four distinct roles: groups that did not interact with other groups at all (*noninteractors*), those that were primarily initiating interactions with other groups (*information seekers*), those that were primarily engaging in noninitiated interactions with other groups (*responders*), and those that engaged in both initiated and non-initiated intergroup interactions (*mutual interactors*). We discuss the results of this cluster analysis in more detail in Appendix A.

Accordingly, we labeled groups as noninteractors if they had exactly zero normalized instrength and exactly zero normalized outstrength. We labeled groups as information seekers if they had exactly zero normalized instrength and nonzero normalized outstrength. We labeled groups as responders if they had exactly zero normalized outstrength but nonzero normalized instrength. Finally, we labeled groups as mutual interactors if they had both nonzero normalized instrength and nonzero normalized outstrength.

To understand the nature of each social role's interactions in the lab, we evaluated the average durations of the interactions for each social role. While all measures we have described so far sum together a group's number or duration of interactions throughout the lab period, measures

related to separate interactions unfold the subtleties in how these groups behave. We calculated two different measures of the average duration of individual interactions: one for initiated interactions and one for noninitiated interactions. That is, for all of the groups in each role, we calculated the sum of all outstrengths (instrengths) divided by the number of interactions that the groups did (not) initiate. We also determined the average number of individual interactions in which groups within each role engaged, again separated by initiated and noninitiated interactions. These measures were computed across groups in each social role and provided more detail about how groups in each role engaged in intergroup interactions (e.g., whether groups engaged in many short interactions, a few long interactions, etc.).

Finally, we investigated the types of lab groups taking on each social role. Considering all lab groups, we performed chi-square tests to determine if role distributions differed by group gender composition or lab curriculum. We also conducted a preliminary analysis to compare the distributions of all-female, all-male, and mixed-gender groups taking on each role *within* each lab curriculum.

III. RESULTS

In this section, we present our results in two parts according to our two research questions. First, we describe our findings about the evolution of intergroup networks over the course of the semester in each course. Second, we present our findings related to the social roles lab groups take on and the relationship between social roles, lab curriculum, and group gender composition.

A. Intergroup networks

Figure 2 shows example network diagrams for four sessions of two different lab sections (1 and 4). These two sections were chosen to illustrate two different lab curricula (early, traditional and recent, reformed) and two different classroom layouts (see Fig. 1). Each of the four chosen lab sessions coincide with different lab activities and depict the salient trends of each section over time. In the diagrams, each node represents a lab group (two to four students) and is positioned according to the actual table layout in the classroom (see Fig. 1). Lab groups changed between sessions, so the same position of a node over time does not necessarily indicate the behaviors of the same students. Network diagrams for all coded sessions of all five sections can be found in Appendix B.

The diagrams in Fig. 2 suggest general qualitative patterns of networks over time. For example, section 4 exhibits few intergroup interactions (fewer total edges) at both the beginning and the end of the semester, but many intergroup interactions (many total edges) during intermediate sessions. This trend is as we might expect for reformed labs. At the beginning of the semester, students are still becoming familiar with the lab's objectives and their peers, possibly making them hesitant to reach out to other groups. As the semester progresses, lab groups become more familiar with each other and engage in more open-ended investigations where they are encouraged to share ideas with other groups. At the end of the semester, students complete a project lab where each group explores a more individualized topic. It is unusual for multiple lab groups to pursue similar investigations during the project

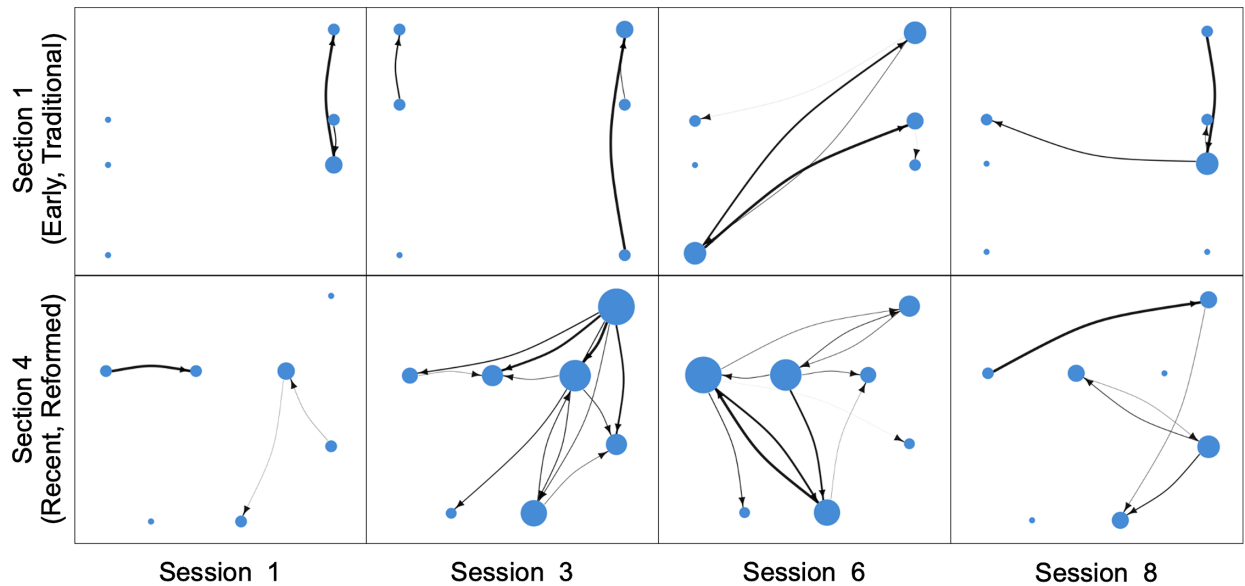


FIG. 2. Sample network diagrams for four sessions of lab sections 1 and 4 produced from our video coding. Each node represents a lab group and nodes are positioned according to the actual table layout of the classroom (Fig. 1). Node sizes are proportional to degree and are scaled the same for all diagrams. Edge thickness is proportional to strength, the sum of the time durations of interactions. Edge arrows point from the group that initiated the interaction to the responding group.

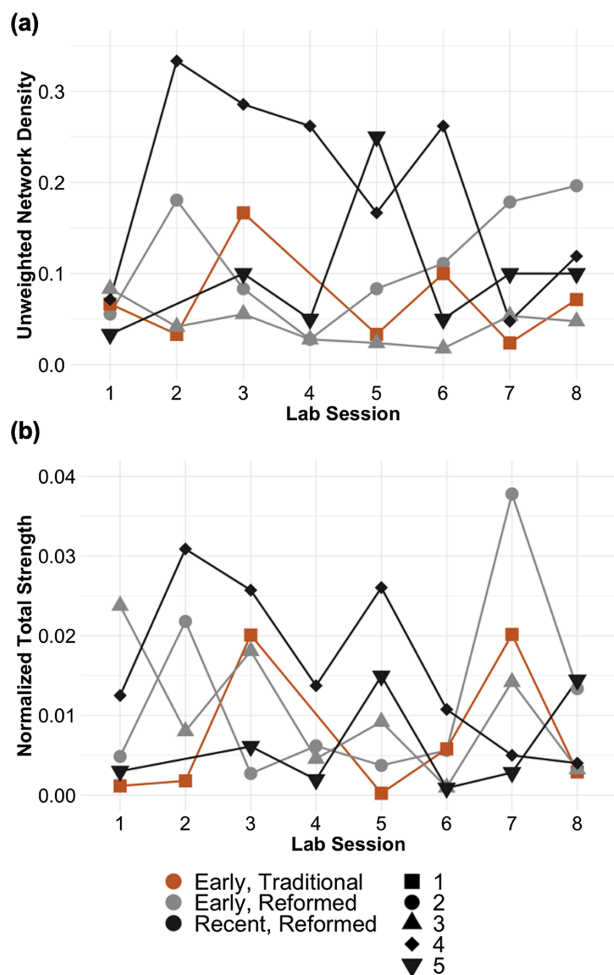


FIG. 3. (a) Unweighted network density, the proportion of observed edges to possible edges between groups, plotted over time for all coded lab sessions. (b) Normalized total strength, the sum of durations of all observed intergroup interactions scaled by the length of the lab session and number of groups present, over time for all coded lab sessions. Colors indicate the lab curriculum and shapes indicate the lab section.

lab, which provides less advantage to interacting with other groups.

In comparison, section 1 (the traditional lab section) exhibits visibly fewer interactions overall (few total edges). Though the traditional lab activities are similar in nature throughout the semester, we still see that the networks change slightly over time. As with section 4, students in the traditional lab are still becoming familiar with the lab and their new peers at the beginning of the semester, evidenced by the sparseness of the session 1 network. However, a few more groups interact with one another in subsequent sessions, perhaps due to increased familiarity with and comfort around their lab peers.

While the diagrams offer useful visualizations of intergroup interaction patterns, quantitative measures across all five sections facilitate broader interpretations. Figure 3(a) shows the unweighted network density—the fraction of

possible edges observed, not taking into account the duration of interactions—of all coded sessions over time. Patterns in the plot [Fig. 3(a)] align with the patterns visible in the sample network diagrams (Fig. 2). For example, section 4's unweighted network density is low during the first session, remains quite high during the middle sessions, and then decreases for sessions 7 and 8 (the project lab).

The most apparent trend in Fig. 3(a), however, is its overall noisiness: the unweighted densities are highly variable across different lab sessions and sections. There is no common trend over time that all (or subsets of) sections follow. While section 4's density is generally higher than all other sections, no section's unweighted densities remain consistently higher or lower than the others. We see that the traditional lab section (section 1) has fairly low unweighted densities throughout, other than a spike in session 3. One of the early, reformed labs (section 3) also has low unweighted densities throughout, while the other early, reformed lab section (section 2) has low to medium unweighted densities (with spikes in sessions 2, 7, and 8). Finally, the other recent, reformed lab (section 5), unlike section 4, has moderate unweighted density values throughout with a dramatic spike in session 5. In this fifth session, there were only five lab groups present and one group went around to every other group asking about their findings, amplifying the unweighted density value. The corresponding network diagram can be found in Appendix B.

That there are no consistent trends across sections or between lab curricula is particularly interesting given that some sections share many features and still do not have aligning trajectories. For instance, students in both the early and recent reformed labs (sections 2 through 5) completed very similar activities during each lab session (see Table III). Nonetheless, all four of these sections exhibit varying network densities within each session [shown in light and dark gray in Fig. 3(a)]. Further, sections 4 and 5 took place within the same semester, course, lab curriculum, and classroom and were instructed by the same TA. Despite these similarities, we see that these two sections exhibit remarkably different unweighted densities during any given session. Such variability within sessions suggests that the proportion of lab groups that interact with one another is not dependent on the lab activity itself nor the course, classroom layout, or TA.

Rather, the prevalence of intergroup interactions may be due to other attributes of a lab section. We note two key differences between sections 4 and 5 that may account for the different numbers of observed interactions: the section populations and the time of day of the sections. As summarized in Table II, section 4 is majority men and section 5 is majority women. Section 4 is also noticeably larger in total enrollment (19 students) than section 5 (11 students). These differences suggest that the student population of a lab section may impact intergroup interaction

patterns. In addition, section 4 took place in the afternoon whereas section 5 took place at night. We suspect not only that students are less prone to interact during labs that take place in the evening based on the sheer time of day, but also that the population of students who enroll in an evening section may be inherently different than that of a section during the day (e.g., student athletes or students with part-time jobs may necessarily enroll in evening sections).

Figure 3(b) shows the normalized total strength—the average fraction of the total lab time each group spent interacting with other groups—for each lab session over time. Similar to Fig. 3(a), the normalized total strength across sections and over time is quite variable. We again see that even among sections with the same lab curriculum (sections 2 and 3 and sections 4 and 5) or the same TA (sections 4 and 5), there is no alignment in the values and trajectories of this network measure. That is, the values of normalized total strength vary widely among lab sections for the same session. This suggests that the proportion of lab time spent on intergroup interactions, in addition to the proportion of lab groups who interact with one another (unweighted network density), is independent of the specific lab activity.

Instead, the normalized total strength of a lab section seems to be dependent on the lab curriculum—traditional or reformed—and the time of day of the section. The normalized total strengths of both the traditional lab section [section 1, shown in red in Fig. 3(b)] and the recent, reformed lab section that took place in the evening [section 5, shown in dark gray in Fig. 3(b)] are typically lower than that of the other sections. This suggests that groups in the traditional and evening lab sections spend less time overall engaging in intergroup interactions. Viewing interactions as the sharing of information, we infer that lab groups in traditional and/or evening labs may be less prone than lab groups in reformed or daytime labs to exchange substantive information with other groups.

Though three of the reformed labs have higher normalized total strengths than the traditional lab, each reformed lab section again exhibits a unique trajectory over time. Combining the information from both unweighted network density and normalized total strength, we also observe variable interaction patterns across these reformed sections. For example, section 4's normalized total strength is quite high throughout the semester but tapers off toward the end. Looking at both Figs. 3(a) and 3(b), session 6 of section 4 contained many interactions but these were relatively short in length. We can see this in the many thin edges of this session's network in Fig. 2. In contrast, section 2's normalized total strength is generally low, but has a large pickup at sessions 2 and 7. The duration of interactions is quite large relative to the unweighted network density in session 7 (for example, as compared to session 2 for section 4), indicating fewer but longer interactions. Broader interpretations of this variability will be explored in the discussion.

B. Group-level social roles

To understand group-level behaviors, we classified groups taking on different social roles and examined what group attributes may impact the role they ultimately take on. Figure 4 shows a point for every observed lab group using the two continuous centrality variables we chose as a basis for role classification: normalized instrength and normalized outstrength. Because both centrality measures are normalized as a fraction of the duration of the lab session, possible values range from 0 to 1 with higher values indicating longer time spent interacting. These two variables exhibit power law distributions as is typical for network data [61], making it difficult to tease apart the many groups having zero or low centrality measures. As a result, we performed a fourth root transformation on both measures (as shown in Fig. 4) to normalize the raw data and find patterns that would otherwise be difficult to resolve. We chose a power law transformation over other types of transformations, such as logarithmic transformations, because these others would inherently change the structure of our observed networks. We discuss this data transformation in more detail in Appendix A.

Table VI enumerates the role classification results, including each role's mean centrality measures and mean durations and numbers of both initiated and noninitiated interactions. We see in Fig. 4 that our role classification scheme meaningfully deciphers groups by their social behavior:

Noninteractors: Lab groups labeled as noninteractors do not interact with other groups at all, as every group taking on this role has zero normalized instrength and zero normalized outstrength. About 35% of the lab groups observed in this study were classified under this role (shown in blue in Fig. 4).

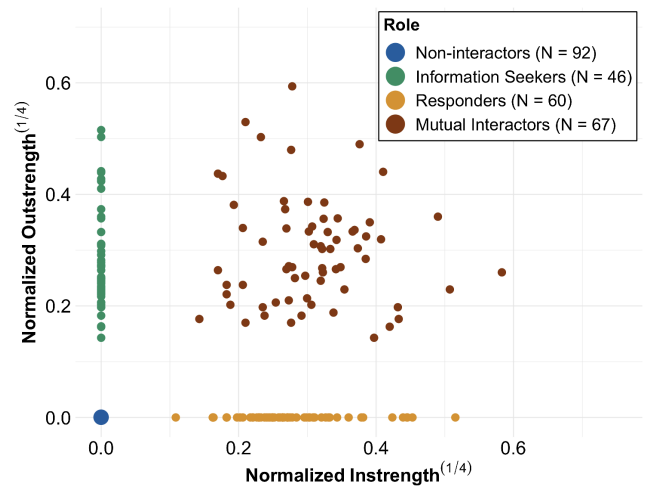


FIG. 4. Transformed scatterplot with a point for each observed lab group based on normalized outstrength and normalized instrength. Colors of each point indicate the social role assigned to the group by our analysis.

TABLE VI. Summary of social role analysis including each role's mean centrality measures, mean durations and numbers of interactions, and number of groups of each gender composition and lab curriculum.

	Noninteractors	Information seekers	Responders	Mutual interactors
Mean normalized instrength	0	0	0.009	0.014
Mean normalized outstrength	0	0.012	0	0.014
Mean \pm SD duration of initiated interactions	0 sec	77 ± 15 sec	0 sec	97 ± 18 sec
Mean \pm SD number of initiated interactions	0	2 ± 0.4	0	2.8 ± 0.3
Mean \pm SD duration of noninitiated interactions	0 sec	0 sec	62 ± 10 sec	95 ± 15 sec
Mean \pm SD number of noninitiated interactions	0	0	1.8 ± 0.2	2.6 ± 0.3
% (N) of total groups	35% (92)	17% (46)	23% (60)	25% (67)
% (N) of all-female groups	35% (9)	11% (3)	27% (7)	27% (7)
% (N) of all-male groups	36% (45)	16% (20)	17% (21)	31% (38)
% (N) of mixed-gender groups	33% (38)	20% (23)	28% (32)	19% (22)
% (N) of Early, Traditional lab groups	48% (19)	18% (7)	23% (9)	11% (5)
% (N) of Early, Reformed lab groups	38% (51)	18% (24)	17% (23)	27% (36)
% (N) of Recent, Reformed lab groups	24% (22)	16% (15)	31% (28)	29% (26)

Information seekers: Lab groups labeled as information seekers only engage in intergroup interactions that they initiate, as all groups in this role have zero normalized instrength and nonzero normalized outstrength. Information-seeking groups on average spend 77 sec per initiated interaction and engage in two separate initiated interactions. About 17% of the lab groups observed in this study were classified under this role (shown in green in Fig. 4).

Responders: Lab groups labeled as responders only engage in intergroup interactions that other groups initiate, as all groups in this role have nonzero normalized instrength and zero normalized outstrength. Responding groups on average spend 62 sec per noninitiated interaction and engage in two separate noninitiated interactions. About 23% of the lab groups observed in this study were classified under this role (shown in orange in Fig. 4).

Mutual interactors: Lab groups labeled as mutual interactors engage in both initiated and noninitiated intergroup interactions, as all groups in this role have nonzero normalized instrength and nonzero normalized outstrength. Mutual interactors on average spend 97 sec per initiated interaction and engage in three separate initiated interactions. These groups on average spend 95 sec per noninitiated interaction and engage in three separate noninitiated interactions. About 25% of the lab groups observed in this study were classified under this role (shown in maroon in Fig. 4).

To illustrate this social role classification for one session, Fig. 5 plots the network diagram for session 2 of section 4. We chose this session because it contained the most intergroup interactions and it exemplifies well three of the social roles we identified. In this session, one group is an information seeker (shown in green), one group is a responder (shown in orange), and the remaining five groups are mutual interactors (shown in maroon). The information seeker group initiates three interactions (three edges pointing away from the green node) and does not engage in any

noninitiated interactions. The responder group exclusively engages in noninitiated interactions (two edges pointing toward the orange node). The remaining five groups all have both incoming and outgoing edges, indicating that they engaged in both initiated and noninitiated intergroup interactions. No groups in this session were noninteractors, however, this role is easily understood as nodes having zero attached edges in the network.

Table VI also summarizes the types of groups that take on each role. We see that all-female, all-male, and mixed-gender groups take on the noninteractor role a similar

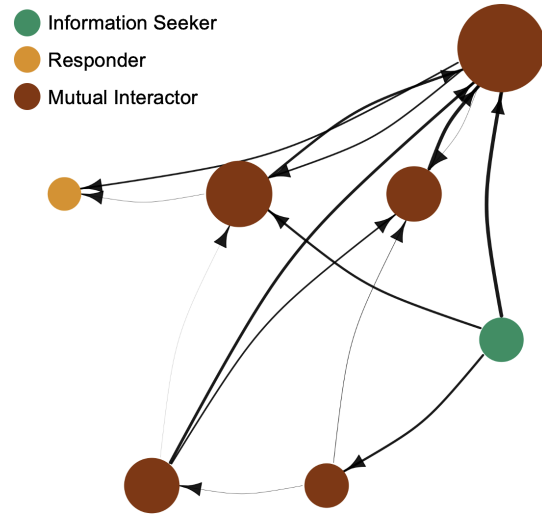


FIG. 5. Network diagram of session 2 of section 4. Each node represents a lab group and is positioned according to the actual table layout of the classroom (Fig. 1). Node size is proportional to degree. Edge thickness is proportional to strength. Edge arrows point from the group that initiated the interaction to the responding group. Nodes are colored by the social role assigned to them by our analysis.

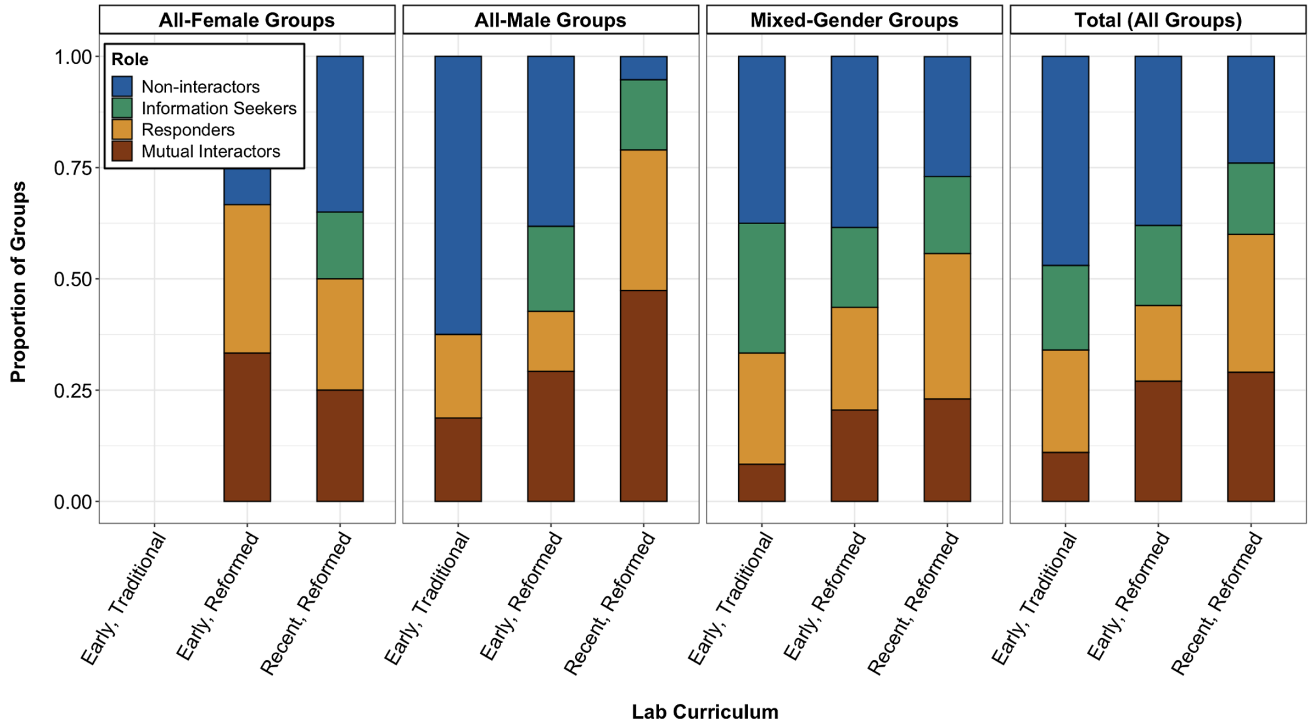


FIG. 6. Stacked bar charts of the proportion of groups of each gender composition and in total that take on each social role within each lab curriculum. The early, traditional lab contained no all-female groups.

amount (between 33% and 36%), with somewhat higher variability in the other roles. Congruently, a chi-square test reveals that there is no statistically significant association between social roles and group gender composition [$\chi^2(6) = 7.8$, $p = 0.25$].

However, the association between social roles and lab curriculum is significant [$\chi^2(6) = 13.3$, $p = 0.038$] with a moderate effect size (Cramer's $V = 0.16$, 95% C.I. = [0.11, 0.26]). As summarized in the last three rows of Table VI, lab groups in the early, traditional lab disproportionately act as noninteractors (48%), with very few groups in this lab taking on the mutual interactor role (11%). The most common role for groups in the early, reformed labs is noninteractor as well (38%), though over a quarter of groups (27%) act as mutual interactors. Groups in the recent, reformed labs are primarily responders (31%) and mutual interactors (29%), with comparably few noninteracting groups (24%). That is, most traditional lab groups are completely isolated and noninteracting (as can be seen in Figs. 2 and 15 in Appendix B) whereas most lab groups (76%) in the recent, reformed labs engage in at least one intergroup interaction.

We further broke down the distribution of social roles by group gender composition *within* each lab curriculum, as shown in Fig. 6. Because of the low number of all-female groups represented in this study, it is difficult to make claims about this gender composition. At a general level, we see that all-female groups do not act as information seekers in the early, reformed lab, but do take on this role in

the recent, reformed lab. In terms of the other three roles, the distribution that all-female groups exhibit is quite similar across the early and recent reformed labs.

All-male groups are more interactive in the recent, reformed lab (only 5% are noninteractors) than in either the traditional (more than 50% noninteractors) or early, reformed lab (more than 30% noninteractors). In parallel, all-male groups are disproportionately mutual interactors in the recent, reformed labs (47%)—noticeably more than either the all-female (25%) or mixed-gender groups (23%) in those labs. We also observe that all-male groups never act as information seekers in the traditional lab, while all-male groups take on this role in both renditions of the reformed labs.

The most noticeable difference for mixed-gender groups is that these groups act as mutual interactors more frequently in the early and recent reformed labs (between 20% and 25%) than in the traditional lab (8%). This increase in mutual interactors is countered by a decrease in the proportion of mixed-gender groups acting as information seekers (about 30% in the traditional lab and about 17% in both early and recent reformed labs). We discuss the implications of this social role breakdown by both gender composition and lab curriculum in the next section.

IV. DISCUSSION

In this study, we coded video recordings of five lab sections throughout a whole semester for the existence and duration of social interactions between lab groups. Though

most prior work on social dynamics in introductory physics labs focuses on within-group interactions [21–27], we aimed to understand the extent and patterns of interactions *between* lab groups. We specifically examined how intergroup lab networks evolve over time for different lab sections and identified the group-level social roles that lab groups take on.

In response to our first research question, we found from our analysis of unweighted network density that the connectedness of lab groups is not noticeably or regularly different across lab curricula. That is, although the nature of the lab tasks in the traditional and reformed labs were quite different, the *number* of intergroup interactions (taking into account direction of interactions) was similar in both settings. We emphasize that the measure of unweighted network density does not take into account the duration or number of separate interactions, rather it captures the fraction of possible edges we observed (i.e., the fraction of groups that interacted with each other at all).

In contrast, we found that the normalized total strength of lab networks was generally higher for three of the four reformed labs than the traditional lab. That is, students spent a larger fraction of the lab session engaging in intergroup interactions in these reformed labs than in the traditional lab and the remaining reformed lab. With the results of the unweighted network density, this suggests that students in the reformed labs engaged in many prolonged intergroup interactions, while students in the traditional lab engaged in a few brief intergroup interactions. These findings align with those of biology education researchers [45,46] and suggest that the nature of open-ended physics labs is likely conducive to more substantive conversations between lab groups than the single-right-answer and single-procedure tasks in traditional labs. Reformed physics labs may offer students more opportunities to develop their experimentation skills by troubleshooting with other groups [28]. Our results also indicate that reformed labs tend more strongly to the communicative goals put forth by the American Association of Physics Teachers [30] than traditional labs.

Yet, the reformed lab section whose normalized total strength was as low as the traditional lab points to another lab section attribute that may impact the amount of intergroup interaction: the time of day of the section. As mentioned, this particular reformed section occurred during the evening, while the other four observed sections took place during the morning or afternoon, and the instructor for this section also led section 4 (which had some of the most interactions). Our results indicate that, despite the ability of the lab instructions and instructor to promote intergroup interactions, night labs may not be as conducive to intergroup interactions as lab sections occurring during the daytime. This may be solely due to the time of day, but we also suspect that the types of students that enroll in night labs (e.g., student athletes and students holding part-time jobs) may play a role in the level of interactivity. Because

the instructor was the same in sections 4 and 5, we can rule out the possibility of interactions being a result of instructor characteristics. We note that these findings should be taken as preliminary, as we only analyzed one traditional lab section and one evening lab section. Future work analyzing the intergroup interactions in more traditional lab sections and sections at different times of day should be conducted to test our baseline claims.

A closer examination of lab sessions with comparably high interactions also offers insight into how instructors may promote intergroup interactions in labs. We discuss here one particular session (session 2 of section 4) of the recent, reformed labs that had the highest unweighted network density and second highest normalized total strength. In this session, students investigated the angular dependence of a pendulum (see Ref. [52]). Only one group in the session performed an experiment precise enough to distinguish the periods when released from 10° and 20° , while the other six groups did not. The instructor explicitly prompted the class to talk to other groups if they were not convinced of this distinguishability, after which most of the lab groups flocked to the one group who obtained this result to ask about their experimental procedure. Our quantitative analysis combined with these episode details suggest that interactions among different lab groups are facilitated by instructor encouragement to do so [62] and that instructors can intentionally specify which groups are productive ones with whom to interact. However, while evidence from this one session demonstrates that students do follow such instructor prompts, the number and duration of interactions in this section mostly decreased for the remainder of the semester. This pattern provides an important addition to our instructional implications: while explicit instructor prompts can encourage intergroup interactions, these prompts likely need to be repeated throughout the semester.

In addressing our second research question, we identified four social roles that lab groups take on: noninteractors, information seekers, responders, and mutual interactors. Noninteractors engage in no intergroup interactions, while mutual interactors engage in both initiated and noninitiated interactions. Moreover, mutual interactors engage in roughly six separate intergroup interactions per session (three initiated intergroup interactions and three noninitiated intergroup interactions). The remaining two roles are dichotomous: information seekers exclusively engage in initiated intergroup interactions and responders exclusively engage in noninitiated intergroup interactions. Both information seekers and responders engage in two separate interactions per session, fewer than mutual interactors. Nearly half (42%) of our observed lab groups actively chose to engage in intergroup interactions (i.e., took on the information seeker or mutual interactor roles), evidence that students not only value socialization in their introductory physics labs [19,20], but actually enact these values in the classroom.

We found that the distribution of lab groups that took on each of these four social roles was significantly different

between lab curricula. Groups in the traditional lab disproportionately (48%) took on the noninteractor role, while the lab groups in the two reformed lab curricula were more evenly spread among the other three roles. Previously, we had found that the unweighted density (the fraction of possible edges present) was similar between the traditional and reformed labs. The proportion of noninteractor groups, however, adds nuance to this result. Because our networks are directed, up to two edges can appear between a pair of groups based on which of the groups initiated each interaction. Thus, the unweighted densities of two networks can be comparable if a few groups are connected with two edges in one case and more groups are connected with a single edge in another. The larger number of noninteractor groups in the traditional lab indicates that the intergroup interactions in the traditional lab likely only occurred between a small fraction of the groups present, with many nodes isolated. In other words, fewer lab groups participated in intergroup interactions in the traditional lab than in the reformed labs. These findings further corroborate the claim that students in reformed labs are generally more interactive than those in traditional labs [45,46].

Notably, only 24% of lab groups in the recent, reformed labs were noninteractors, compared with 38% of lab groups in the early, reformed labs. These results suggest that the evidence-based modifications made to the reformed lab instructions between the two analyzed course offerings may have had led to more interactions between lab groups. Alternatively, population differences between students in the early and recent reformed labs may explain the varying social roles. Students in the early, reformed labs were less diverse than those in the recent, reformed labs. For example, most students in the early, reformed lab intended to major in physics, were first-semester college students, and were part of majority demographic groups in the discipline (i.e., men of Asian/Asian American or White race or ethnicity). In contrast, most students in the recent, reformed lab intended to major in engineering or other STEM disciplines and were more heterogeneous across demographic variables (gender and race or ethnicity).

To probe the effects of demographics, we evaluated whether groups' gender composition impacted the social role they took on. We found that the social role a lab group took on did not depend on gender composition. Our results also show that no group gender composition disproportionately acted as either responders or information seekers, contrary to our expectations, suggesting that biased peer perceptions were not at play in our analyzed lab sections. If they were, we would expect all-male groups to largely act as responders (other groups coming to them for help) given the documented gender bias in students' perceptions of their peers in STEM courses [48,49]. However, these prior studies were conducted at the course level, largely probing students' peer perceptions in lecture. Our study suggests that future work should investigate whether the same level

of gender bias occurs in the context of lab rather than in the context of lecture or discussion sections.

Finally, parsing social roles by *both* lab curriculum and gender composition showed general trends occurring at the intersection of these two variables. Though the number of all-female groups was low in the study, none of the all-female groups acted as information seekers in the early, reformed lab but a small fraction took on this role in the recent, reformed labs. All-male groups were noticeably more interactive in the reformed labs than in the traditional lab. This change was largely a shift from all-male groups acting as noninteractors to acting as information seekers and mutual interactors. Similarly, more mixed-gender groups acted as mutual interactors in the reformed labs than in the traditional lab. At the same time, fewer mixed-gender groups acted as information seekers in the reformed labs than in the traditional lab. Overall, it seems that the recent, reformed lab offered more opportunities for all group gender compositions to take on the three interactive social roles (information seekers, responders, and mutual interactors) compared with the other lab curricula. Based on limited statistics, these observations should be viewed as preliminary and should be tested in future work.

We end this section by acknowledging a few additional limitations of our study. First, the demographic focus of our analyses has been related to gender. We understand that identity is multifaceted and that other demographic features of individuals, such as race or ethnicity, age, and socioeconomic status, also impact the way students engage in labs. Future research should examine the role of these other factors in intergroup interactions, as well as examine the replicability of these findings with other populations of students, such as at other institutions. Second, our five analyzed lab sections were different across multiple variables, including physics course, classroom layout, instructor, and time of day. This renders our claims about the impact of lab instruction limited and we hope that future work uses similar methods to examine traditional and reformed labs with fewer between-section differences. Finally, the aim of this study was to examine interactions at the group level. We acknowledge that group behaviors may represent or be heavily influenced by the personalities of individual group members (e.g., one outgoing group member initiating many interactions with other groups). While our quantitative methods do not allow for us to meaningfully tease out such tendencies, future work should qualitatively investigate the relationship between individual and group-level interaction patterns.

V. CONCLUSION

Most prior work on social interactions in introductory physics labs exclusively focuses on within-group interactions. However, interacting with peers outside of one's own lab group is beneficial for learning about possible procedures and results, particularly in reformed labs where

students have more control over the experimental designs and outcomes. In this study, we examined the extent and patterns of intergroup interactions in various lab contexts using SNA. Our results from coding video of five different lab sections suggest that lab instruction has the potential to influence the extent to which students engage in intergroup interactions. In particular, we offer preliminary evidence that more prolonged intergroup interactions take place in most of the reformed labs than in a traditional lab, with the caveat that one of our observed reformed lab sections that took place in the evening was comparable to the traditional lab. We also identified four social roles that lab groups take on: noninteractors, information seekers, responders, and mutual interactors. We found that lab curriculum affected social role distributions, such that lab groups in the traditional lab disproportionately acted as noninteractors, while lab groups in reformed labs were more distributed between the other three roles. Our results also present possible, but unclear, effects of group gender composition on the social roles in each lab context.

Our study provides meaningful implications for instructors of introductory physics labs. We saw that direct prompts from the lab instructor to interact with other groups did indeed lead to more and longer interactions. Moreover, this instructor specified exactly which lab group others may benefit from talking to, which stimulated collaboration. These prompts were also driven by the lab task: groups found very different results from one another and in response talked to one another about their experimental procedures. In reformed labs where experimental designs are not directly provided, such instructor moves that trigger students' interactions with other groups may allow for groups to refine their experimental decisions. At the same time, however, we found that these direct prompts toward intergroup interactions given in one session did not carry over to future sessions and that very different amounts of intergroup interactions took place in two lab sections taught by this same instructor. Future work should further examine the role of the instructor in stimulating interactions among lab groups and whether this evolves over time and differs across lab curricula.

The work presented here provides quantitative evidence of intergroup interactions in instructional physics labs, demonstrating the fruitful potential of combining video data and SNA to understand student interaction patterns. Given the implications for instruction, future work should seek to understand the substance and effects of intergroup interactions beyond the numbers and durations. For instance, a qualitative analysis of video of interactions between lab groups might afford insight into the "quality" of such interactions. It could be the case that a short interaction provides meaningful next steps to a group's experiment, while a longer interaction, for example, about students' weekend plans, is entirely off topic. Understanding what students talk about when interacting across lab groups and

what types of problems or misunderstandings drive groups to seek external help will further inform curriculum development and instructional practice.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program No. DGE-1650441 and by Cornell University's College of Arts and Sciences Active Learning Initiative. We thank undergraduate student Daniyar Kushaliev for contributing to the video coding and Eric Brewe, Michelle Smith, and David Esparza for engaging in meaningful discussions about this work.

APPENDIX A: CLUSTER ANALYSIS RESULTS FOR SOCIAL ROLE CLASSIFICATION

For the second part of our main analysis, we aimed to classify the $N = 265$ lab groups observed as part of the study into social roles based on their normalized instrength and normalized outstrength. As is typical with network centrality measures, distributions of both of these dimensions were highly skewed [61]. A cluster analysis of the raw data only meaningfully distinguished the outliers for each centrality measure and did not tease apart the tight distribution of groups with centrality values close to zero. Though numerically all of these tightly packed groups interacted with other groups a similar amount, we sought to pull apart groups who engaged in any amount of intergroup interactions from those who were strictly non-interacting.

We aimed to normalize the data as much as possible before clustering. We could not perform a log-transformation because our centrality distributions contain many zero values, for which the log-transform value is undefined. If we were to shift these undefined values by taking the log-transform of our centrality measures plus some constant,

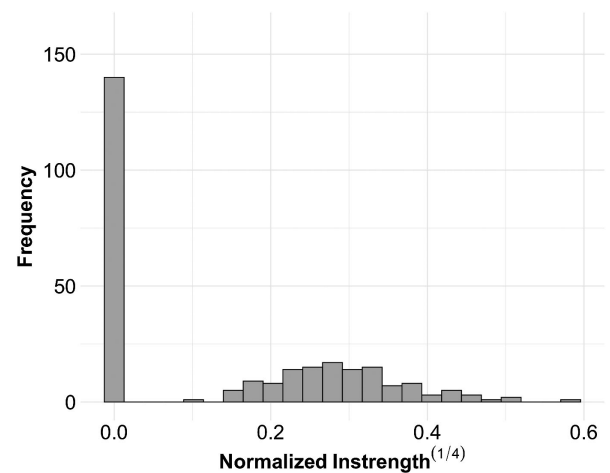


FIG. 7. Histogram of the normalized instrength data using the fourth root transformation. Though we maintain a peak at zero, the rest of the data exhibit a fairly normal distribution.

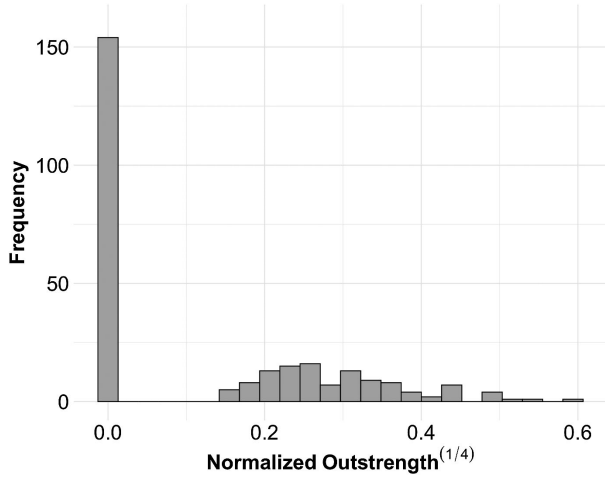


FIG. 8. Histogram of the normalized outstrength data using the fourth root transformation. Though we maintain a peak at zero, the rest of the data exhibit a fairly normal distribution.

say one, we would inherently change the structure of our observed networks by implying that every possible edge exists. This option was thus not suitable for our data. So, we turned to power law transformations and found that the data first start to become normalized (with the exception of a peak at zero) at the fourth root transformation (it remains skewed at the second and third root transformations), as shown in Figs. 7 and 8.

With these transformed data, we ran a k -means clustering algorithm to determine the most meaningful clustering of groups. To decide the optimal number of clusters, we used the elbow method [63]. This method maximizes the proportion of total sum of squares explained by the between-cluster sum of squares. Figure 9 shows the elbow plot for our data, indicating that $k = 4$ clusters is the best choice. The results of the clustering are shown in Fig. 10

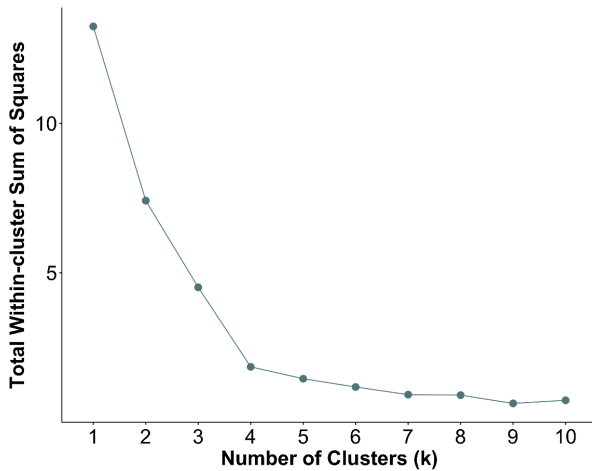


FIG. 9. Elbow plot for the k -means cluster analysis using the data transformed with the fourth root function. Maximizing the proportion of total sum of squares explained by the between-cluster sum of squares informed our choice to perform cluster analysis with $k = 4$.

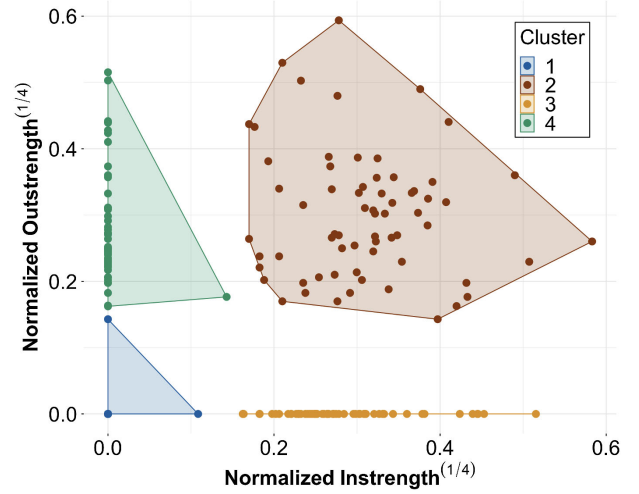


FIG. 10. Cluster plot showing a point for each lab group based on their measures of normalized instrength and normalized outstrength. Point colors indicate the cluster to which the group was assigned using the k -means clustering algorithm.

and informed our social role classification. We see that groups in Cluster 1 are essentially *noninteractors* with small or zero values for both measures. Lab groups in Cluster 2 have nonzero normalized instrength and nonzero normalized outstrength, i.e., they are *mutual interactors*. Cluster 3 groups exclusively engage in noninitiated inter-group interactions (*responders*) and almost all Cluster 4 groups exclusively engage in initiated interactions (*information seekers*). The classification scheme used in the main text almost duplicates this clustering (see Fig. 4), with the exception of a few lab groups that we relabeled in order to maintain a consistent and well-defined classification process (described in Sec. II E.).

Importantly, we see in Fig. 10 that even at this level of data transformation, there are no subtleties or patterns in the data that are overlooked by our role classification. Even within the mutual interactor cluster, there is no structure to the data that suggests it can be broken down into smaller groupings with more specific interaction patterns.

APPENDIX B: NETWORK DIAGRAMS FOR ALL SECTIONS

Figures 11–15 display the network diagrams for all coded sessions of all five lab sections. Each node represents a lab group and nodes are positioned according to the actual table layout of the classroom (see Fig. 1). In some sessions, not all tables were occupied by a lab group, so there is no node present at that table location. Lab groups were not necessarily the same in all sessions, so each node does not represent the same group of students in all diagrams. Node sizes are proportional to degree and are scaled the same for all diagrams. Edge thickness is proportional to total strength, the sum of the time durations of separate interactions. Edge arrows point from the group that initiated the interaction to the responding group.

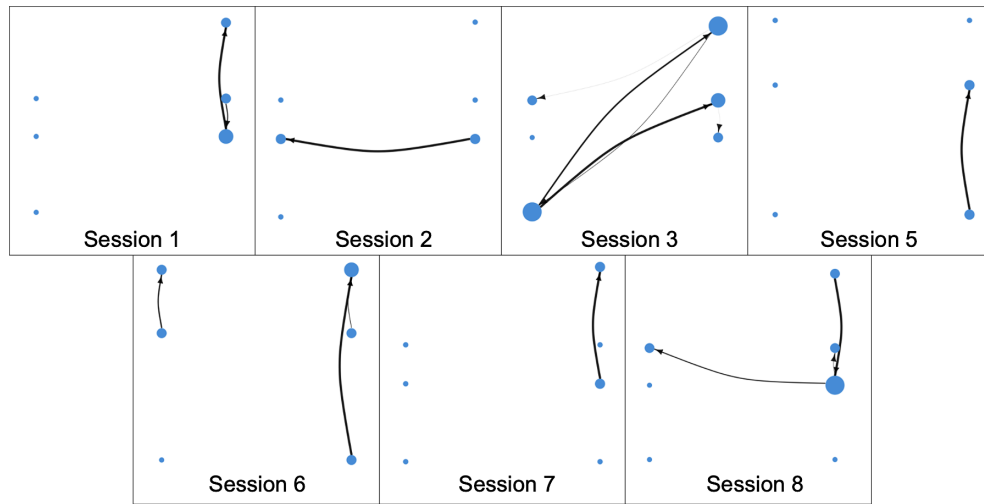


FIG. 11. Network diagrams for all coded sessions of section 1.

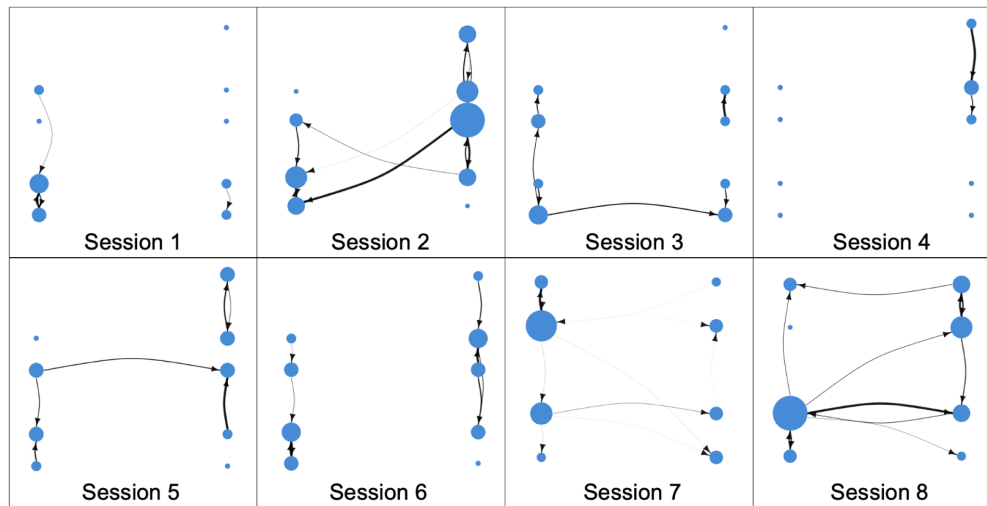


FIG. 12. Network diagrams for all coded sessions of section 2.

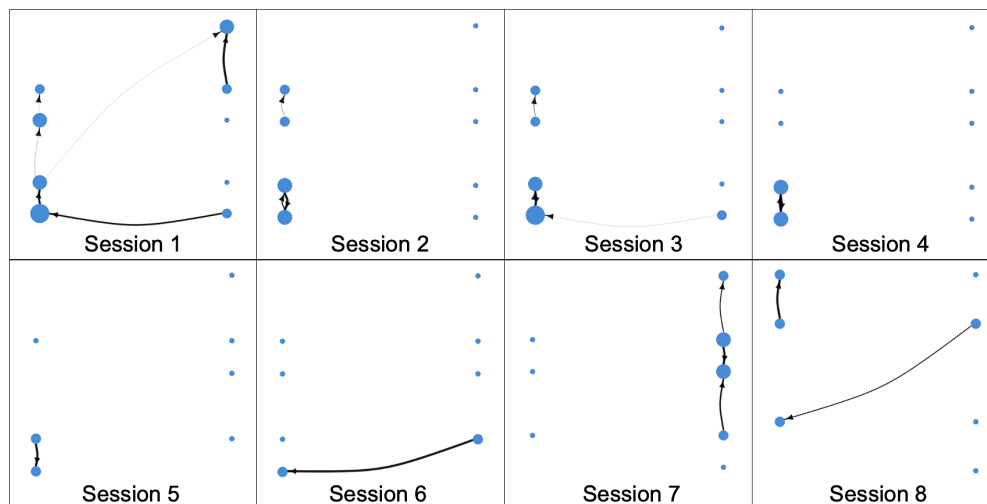


FIG. 13. Network diagrams for all coded sessions of section 3.

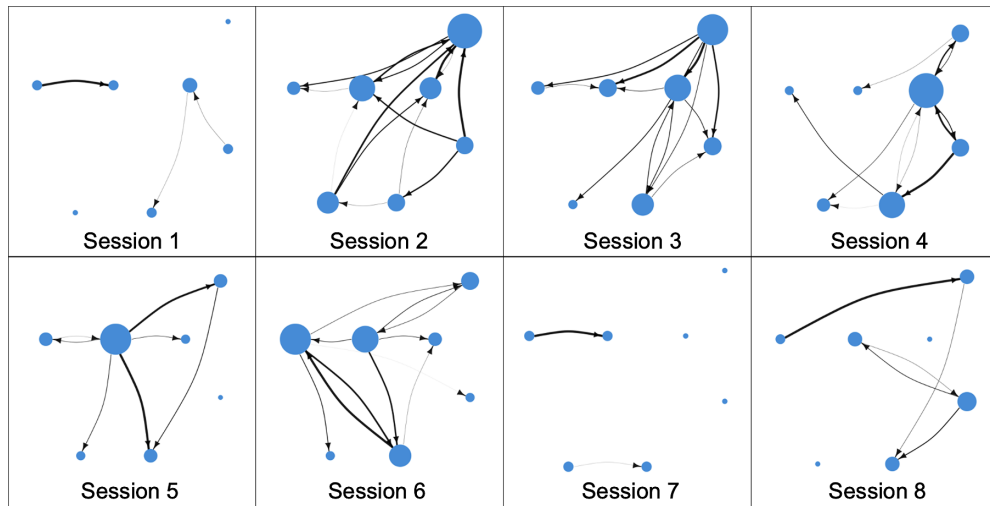


FIG. 14. Network diagrams for all coded sessions of section 4.

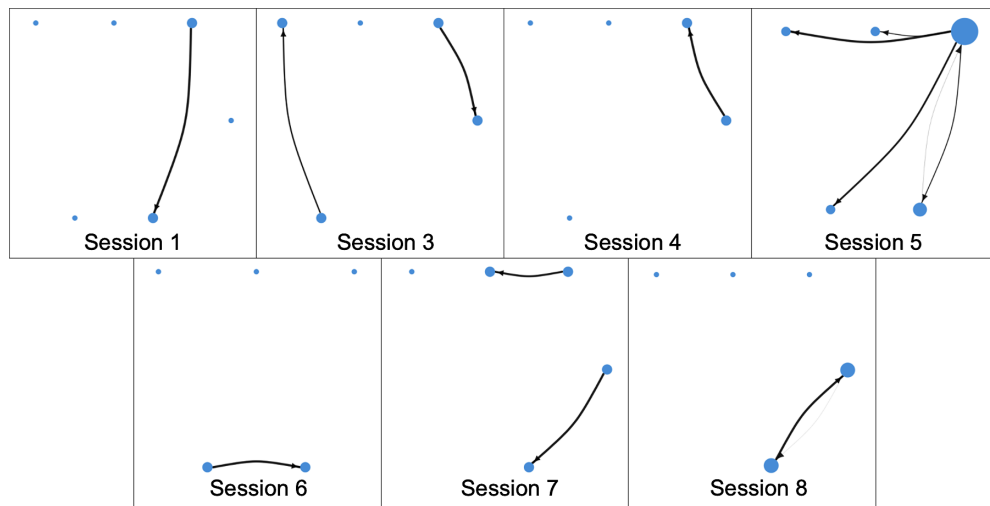


FIG. 15. Network diagrams for all coded sessions of section 5.

-
- [1] J. Piaget and M. Trans Cook, *The Origins of Intelligence in Children* (W. W. Norton & Co., New York, 1952).
 - [2] Lev S. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes*, edited by M. Cole *et al.* (Harvard University Press, Cambridge, Massachusetts, 1978).
 - [3] B. Rogoff, E. Matusov, and C. White, Models of teaching and learning: Participation in a community of learners, in *The Handbook of Education and Human Development* (Blackwell, Oxford, UK, 1996), pp. 388–414.
 - [4] M. Crotty and M.F. Crotty, *The Foundations of Social Research: Meaning and Perspective in the Research Process* (Routledge, London, 1998).
 - [5] J. Pelech and G. Pieper, *The Comprehensive Handbook of Constructivist Teaching: From Theory to Practice* (Information Age Publishing (IAP), Charlotte, NC, 2010).
 - [6] E. Williams, E. Brewe, J.P. Zwolak, and R. Dou, Understanding centrality: investigating student outcomes within a classroom social network, in *Proceedings of the 2015 Physics Education Research Conference, College Park, MD* (AIP, New York, 2015), pp. 375–378.
 - [7] S. Dokuka, D. Valeeva, and M. Yudkevich, How academic achievement spreads: The role of distinct social networks in academic performance diffusion, [PLoS One **15**, e0236737 \(2020\)](https://doi.org/10.1371/journal.pone.0236737).

- [8] J. Bruun and E. Brewé, Talking and learning physics: Predicting future grades from network measures and force concept inventory pretest scores, *Phys. Rev. ST Phys. Educ. Res.* **9**, 020109 (2013).
- [9] A. Traxler, A. Gavrin, and R. Lindell, Networks identify productive forum discussions, *Phys. Rev. Phys. Educ. Res.* **14**, 020107 (2018).
- [10] R. Dou, E. Brewé, J. P. Zwolak, G. Potvin, E. A. Williams, and L. H. Kramer, Beyond performance metrics: Examining a decrease in students physics self-efficacy through a social networks lens, *Phys. Rev. Phys. Educ. Res.* **12**, 020124 (2016).
- [11] J. P. Zwolak, R. Dou, E. A. Williams, and E. Brewé, Students network integration as a predictor of persistence in introductory physics courses, *Phys. Rev. Phys. Educ. Res.* **13**, 010113 (2017).
- [12] J. P. Zwolak, M. Zwolak, and E. Brewé, Educational commitment and social networking: The power of informal networks, *Phys. Rev. Phys. Educ. Res.* **14**, 010131 (2018).
- [13] E. Brewé, L. Kramer, and V. Sawtelle, Investigating student communities with network analysis of interactions in a physics learning center, *Phys. Rev. ST Phys. Educ. Res.* **8**, 010101 (2012).
- [14] R. Dou and J. P. Zwolak, Practitioners guide to social network analysis: Examining physics anxiety in an active-learning setting, *Phys. Rev. Phys. Educ. Res.* **15**, 020105 (2019).
- [15] E. Brewé, L. H. Kramer, and G. E. O'Brien, Changing participation through formation of student learning communities, *AIP Conf. Proc.* **1289**, 85 (2010).
- [16] K. Commeford, E. Brewé, and A. L. Traxler, Characterizing active learning environments in physics: Network analysis of Peer Instruction classroom using ERGMs, in *Proceedings of the 2019 Physics Education Research Conference, Provo, UT* (AIP, New York, 2019), pp. 117–122.
- [17] A. L. Traxler, T. Suda, E. Brewé, and K. Commeford, Network positions in active learning environments in physics, *Phys. Rev. Phys. Educ. Res.* **16**, 020129 (2020).
- [18] E. Brewé, Modeling theory applied: Modeling instruction in introductory physics, *Am. J. Phys.* **76**, 1155 (2008).
- [19] M. Brown and R. M. DeMonbrun, Who gets helped? the opportunity structure of the college physics classroom, peer instruction, and perceptions of help seeking, *J. Coll. Sci. Teach.* **049**, 36 (2019).
- [20] D. J. Rosen and A. M. Kelly, Epistemology, socialization, help seeking, and gender-based views in in-person and online, hands-on undergraduate physics laboratories, *Phys. Rev. Phys. Educ. Res.* **16**, 020116 (2020).
- [21] A. T. Danielsson, Exploring woman university physics students doing gender and doing physics, *Gender Educ.* **24**, 25 (2012).
- [22] K. N. Quinn, M. M. Kelley, K. L. McGill, E. M. Smith, Z. Whipps, and N. G. Holmes, Group roles in unstructured labs show inequitable gender divide, *Phys. Rev. Phys. Educ. Res.* **16**, 010129 (2020).
- [23] A. J. Gonsalves, A. Danielsson, and H. Pettersson, Masculinities and experimental practices in physics: The view from three case studies, *Phys. Rev. Phys. Educ. Res.* **12**, 020120 (2016).
- [24] D. Doucette, R. Clark, and C. Singh, Hermione and the secretary: How gendered task division in introductory physics labs can disrupt equitable learning, *Eur. J. Phys.* **41**, 035702 (2020).
- [25] J. Day, J. B. Stang, N. G. Holmes, D. Kumar, and D. A. Bonn, Gender gaps and gendered action in a first-year physics laboratory, *Phys. Rev. Phys. Educ. Res.* **12**, 020104 (2016).
- [26] J. Jovanovic and S. Steinbach King, Boys and girls in the performance-based science classroom: Who's doing the performing?, *Am. Educ. Res. J.* **35**, 477 (1998).
- [27] N. G. Holmes, G. Heath, K. Hubenig, S. Jeon, Y. Z. Kalender, E. Stump, and E. C. Sayre, Defining and observing physics lab group equity from multiple perspectives (to be published).
- [28] D. Lowe, C. Berry, S. Murray, and E. Lindsay, Adapting a remote laboratory architecture to support collaboration and supervision, *Int. J. Online Biomed. Eng.* **5**, 51 (2009).
- [29] P. Heron and L. E. McNeil, *Phys21: Preparing Physics Students for 21st Century Careers*, Technical report, Joint Task Force on Undergraduate Physics Programs (American Physical Society and American Association of Physics Teachers, College Park, MD, 2016).
- [30] J. Kozminski, H. J. Lewandowski, N. Beverly, S. Lindaas, D. Deardorff, A. Reagan, R. Dietz, R. Tagg, M. Eblen-Zayas, J. Williams, R. Hobbs, and B. M. Zwickl, AAPT committee on laboratories, *AAPT Recommendations for the Undergraduate Physics Laboratory Curriculum* (AAPT, College Park, MD, 2014).
- [31] N. G. Holmes and C. E. Wieman, Introductory physics labs: We can do better, *Phys. Today* **71**, No. 1, 38 (2018).
- [32] E. M. Smith and N. G. Holmes, Best practice for instructional labs, *Nat. Phys.* **17**, 662 (2021).
- [33] E. M. Smith, M. M. Stein, C. Walsh, and N. G. Holmes, Direct measurement of the impact of teaching experimentation in physics labs, *Phys. Rev. X* **10**, 011029 (2020).
- [34] N. G. Holmes, C. E. Wieman, and D. A. Bonn, Teaching critical thinking, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11199 (2015).
- [35] N. G. Holmes, B. Keep, and C. E. Wieman, Developing scientific decision making by structuring and supporting student agency, *Phys. Rev. Phys. Educ. Res.* **16**, 010109 (2020).
- [36] E. Etkina, A. Van Heuvelen, S. White-Brahmia, D. T. Brookes, M. Gentile, S. Murthy, D. Rosengrant, and A. Warren, Scientific abilities and their assessment, *Phys. Rev. ST Phys. Educ. Res.* **2**, 020103 (2006).
- [37] A. Karelina and E. Etkina, Acting like a physicist: Student approach study to experimental design, *Phys. Rev. ST Phys. Educ. Res.* **3**, 020106 (2007).
- [38] E. Etkina, A. Karelina, and M. Ruibal-Villasenor, How long does it take? A study of student acquisition of scientific abilities, *Phys. Rev. ST Phys. Educ. Res.* **4**, 020108 (2008).
- [39] E. Etkina, A. Karelina, M. Ruibal-Villasenor, D. Rosengrant, R. Jordan, and C. E. Hmelo-Silver, Design and reflection help students develop scientific abilities: Learning in introductory physics laboratories, *J. Learn. Sci.* **19**, 54 (2010).

- [40] B. R. Wilcox and H. J. Lewandowski, Open-ended versus guided laboratory activities: Impact on students' beliefs about experimental physics, *Phys. Rev. Phys. Educ. Res.* **12**, 020132 (2016).
- [41] B. R. Wilcox and H. J. Lewandowski, Developing skills versus reinforcing concepts in physics labs: Insight from a survey of students' beliefs about experimental physics, *Phys. Rev. Phys. Educ. Res.* **13**, 010108 (2017).
- [42] N. G. Holmes and H. J. Lewandowski, Investigating the landscape of physics laboratory instruction across North America, *Phys. Rev. Phys. Educ. Res.* **16**, 020162 (2020).
- [43] K. Funkhouser, W. M. Martinez, R. Henderson, and M. D. Caballero, Design, analysis, tools, and apprenticeship (DATA) Lab. *Eur. J. Phys.* **40**, 065701 (2019).
- [44] L. Corwin Auchincloss, S. L. Laursen, J. L. Branchaw, K. Eagan, M. Graham, D. I. Hanauer, G. Lawrie, C. M. McLinn, N. Pelaez, S. Rowland *et al.*, Assessment of course-based undergraduate research experiences: A meeting report, *CBE Life Sci. Educ.* **13**, 29 (2014).
- [45] D. I. Hanauer and G. Hatfull, Measuring networking as an outcome variable in undergraduate research experiences, *CBE Life Sci. Educ.* **14**, ar38 (2015).
- [46] D. Esparza, A. E. Wagler, and J. T. Olimpo, Characterization of instructor and student behaviors in CURE and Non-CURE learning environments: impacts on student motivation, science identity development, and perceptions of the laboratory experience, *CBE Life Sci. Educ.* **19**, ar10 (2020).
- [47] A. Almaatouq, A. Noriega-Campero, A. Alotaibi, P. M. Krafft, M. Moussaid, and A. Pentland, Adaptive social networks promote the wisdom of crowds, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11379 (2020).
- [48] D. Z. Grunspan, S. L. Eddy, S. E. Brownell, B. L. Wiggins, A. J. Crowe, and S. M. Goodreau, Males under-estimate academic performance of their female peers in undergraduate biology classrooms, *PLoS One* **11**, e0148405 (2016).
- [49] B. Bloodhart, M. M. Balgopal, A. M. A. Casper, L. B. Sample McMeeking, and E. V. Fischer, Outperforming yet undervalued: Undergraduate women in STEM, *PLoS One* **15**, e0234685 (2020).
- [50] S. Salehi, N. G. Holmes, and C. Wieman, Exploring bias in mechanical engineering students' perceptions of classmates, *PLoS One* **14**, 0212477 (2019).
- [51] N. G. Holmes and E. M. Smith, Operationalizing the AAPT learning goals for the lab, *Phys. Teach.* **57**, 296 (2019).
- [52] N. G. Holmes and D. A. Bonn, Quantitative comparisons to promote inquiry in the introductory physics lab, *Phys. Teach.* **53**, 352 (2015).
- [53] E. M. Smith and N. G. Holmes, Evaluating instructional labs' use of deliberate practice to teach critical thinking skills, *Phys. Rev. Phys. Educ. Res.* **16**, 020150 (2020).
- [54] M. Sundstrom, A. McLean Phillips, and N. G. Holmes, Problematizing in inquiry-based labs: How students respond to unexpected results, in *Proceedings of the 2020 Physics Education Research Conference*, virtual conference (AIP, New York, 2020).
- [55] A. McLean Phillips, M. Sundstrom, D. G. Wu, and N. G. Holmes, Not engaging with problems in the lab: Students' navigation of conflicting data and models, *Phys. Rev. Phys. Educ. Res.* **17**, 020112 (2021).
- [56] C. J. Walsh, D. Kushaliev, and N. G. Holmes, Connecting the dots: Student social networks in introductory physics labs, in *Proceedings of the 2020 Physics Education Research Conference*, virtual conference (AIP, New York, 2020).
- [57] O. Friard and M. Gamba, Boris: A free, versatile open-source event-logging software for video and audio coding and live observations, *Methods Ecol. Evol.* **7**, 1325 (2016).
- [58] D. Z. Grunspan, B. L. Wiggins, and S. M. Goodreau, Understanding classrooms through social network analysis: A primer for social network analysis in education research, *CBE Life Sci. Educ.* **13**, 167 (2014).
- [59] E. Brewe, The roles of engagement: Network analysis in physics education research, in *Getting Started in PER*, edited by C. Henderson and K. A. Harper (American Association of Physics Teachers, College Park, MD, 2018), Reviews in PER, Vol. 2.
- [60] A. Likas, N. Vlassis, and J. J. Verbeek, The global k-means clustering algorithm, *Pattern Recognit.* **36**, 451 (2003).
- [61] R. Toivonen, J.-P. Onnela, J. Saramäki, J. Hyvönen, and K. Kaski, A model for social networks, *Physica A (Amsterdam)* **371**, 851 (2006).
- [62] D. Winter, P. Lemons, J. Bookman, and W. Hoese, Novice instructors and student-centered instruction: Identifying and addressing obstacles to learning in the college science, *J. Scholarship Teach. Learn.* **2**, 14 (2012); Retrieved from <https://scholarworks.iu.edu/journals/index.php/josotl/article/view/1590>.
- [63] T. M. Kodinariya and P. R. Makwana, Review on determining number of cluster in k-means clustering, *Int. J. Adv. Res. Comput. Sci. Manage. Stud.* **1**, 90 (2013).