Utilizing network analysis to explore student qualitative inferential reasoning chains

J. Caleb Speirs
Department of Physics, University of North Florida, Jacksonville, Florida 32224-7699, USA

MacKenzie R. Stetzer
Department of Physics and Astronomy and Maine Center for Research in STEM Education, University of Maine, Orono, Maine 04469, USA

Beth A. Lindsey
Department of Physics, Penn State Greater Allegheny, McKeesport, Pennsylvania 15132, USA

(Received 1 February 2023; revised 7 November 2023; accepted 26 April 2024; published 29 May 2024)

Over the course of the introductory calculus-based physics course, students are often expected to build conceptual understanding and develop and refine skills in problem solving and qualitative inferential reasoning. Many of the research-based materials developed over the past 30 years by the physics education research community use sequences of scaffolded questions to step students through a qualitative inferential reasoning chain. It is often tacitly assumed that, in addition to building conceptual understanding, such materials improve qualitative reasoning skills. However, clear documentation of the impact of such materials on qualitative reasoning skills is critical. New methodologies are needed to better study reasoning processes and to disentangle, to the extent possible, processes related to physics content from processes general to all human reasoning. As a result, we have employed network analysis methodologies to examine student responses to reasoning-related tasks in order to gain deeper insight into the nature of student reasoning in physics. In this paper, we show that network analysis metrics are both interpretable and valuable when applied to student reasoning data generated from *reasoning chain construction tasks*. We also demonstrate that documentation of improvements in the articulation of specific lines of reasoning can be obtained from a network analysis of responses to reasoning chain construction tasks.

DOI: 10.1103/PhysRevPhysEducRes.20.010147

I. INTRODUCTION

Students pursuing undergraduate science, technology, engineering, and mathematics (STEM) majors outside of physics are often expected to take one or more physics courses as part of their degree programs. While certain physics concepts and principles will be of use in these students' future academic careers, many will not. Instead, it is often expected that the lasting long-term learning outcomes from a physics course will be a repertoire of problem-solving strategies, a familiarity with mathematizing real-world situations, and a strong set of critical thinking skills related to qualitative inferential reasoning. Furthermore, these takeaways are important to all students taking a physics course, including those who go on to be physics majors and physicists.

Physics education research (PER) has produced many instructional materials that have been demonstrated to improve conceptual understanding and to produce other important learning outcomes [1–5]. Many of these materials are scaffolded and step students through qualitative chains of inferences via a series of questions [6–8]. It is often tacitly assumed that such materials also improve qualitative reasoning skills, but there is little documentation of such improvements in the PER literature. Furthermore, it has been observed that despite overall conceptual gains after research-based instruction, there are still certain physics questions for which it is difficult to improve student performance [9–11]. These studies suggest that reasoning processes general to all humans may impact how students understand and reason in a physics context.

There is thus a need to investigate how students generate qualitative inferential chains of reasoning. Many studies have investigated student reasoning in the context of specific physics problems. Some detailed specific reasoning difficulties [12] and some attempted to model the dynamics of student reasoning [13]. Other studies sought to identify domain-general reasoning phenomena (such as heuristics and biases common to all human reasoning [14]) and study their impact on physics-specific reasoning [10,11,15–18]. The methodologies used in these studies are powerful, and a methodology that systematically examines the structure of student reasoning would advance the broader research agenda outlined in this literature—especially if that methodology could separate, to some extent, the structure of the reasoning required on a given problem from the conceptual knowledge that underpins that reasoning.

The paired question methodology reported in Ref. [10] comes close to the goal of separating reasoning skills from conceptual understanding of a given problem. This methodology has been used extensively elsewhere [18–20], particularly in situations that require reasoners to productively navigate an intuitive response when it is in conflict with the correct response. Paired questions have provided further evidence that many students possess an ability to reason correctly through a physics problem but opt for other, more salient lines of reasoning on specific, closely related questions.

More recently, we have developed a new methodology centered on *reasoning chain construction tasks,* or *chaining tasks*, that has also been designed to separate reasoning skills from understanding particular physics concepts. This methodology was initially reported in Ref. [21] and has since been used to leverage results from cognitive science to improve student performance on qualitative physics questions [18]. These tasks have also been implemented in chemistry courses, in which they have been shown to provide deeper insights into student thinking than traditional free-response questions [22]. In this paper, we describe a method for exploring chaining task data using network analysis and present two examples that demonstrate the utility of network analysis methods for gaining insight into the structure of student reasoning via chaining tasks. In combination with reasoning chain construction tasks, network analysis generates novel data related to the content and structure of student arguments. The overarching goal of this manuscript is to describe and highlight the affordances of this novel data.

II. BACKGROUND

In this section, we review pertinent literature that demonstrates the need for more sophisticated analyses of student reasoning and highlights the unique affordances of network analysis of chaining task data to meet this need.

A. Research directly related to qualitative inferential reasoning in physics education

Qualitative inferential reasoning is any type of reasoning that is qualitative in nature and makes inferences, whether deductive or inductive, from given premises. It is common in physics instruction. For instance, consider the following physics question: “An object is at rest on a rough surface but is subject to a forward force of 30 N. What is the value of the friction force from the surface on the object?” A typical way of presenting the solution to this problem is to start with the premise (often called a first principle) that Newton’s second law holds (premise 1). Then, from this premise, one infers that because the object is at rest, all the forces on the object sum to zero (qualitative inference 1). One could then reason that since the only two forces acting horizontally on the object are friction and the applied 30 N force (premise 2), the friction force must therefore be equal to 30 N (qualitative inference 2). In qualitative inferential reasoning, premises (whether or not they are explicitly called such) are combined to make inferences, usually in sequence, about the problem. Such a sequence of inferences could be called a *chain* of inferences. This type of reasoning is distinct from the algebraic manipulations or estimation practices required in some problem-solving activities. However, qualitative inferential reasoning is closely intertwined with the idea of conceptual understanding—after all, it is difficult to measure robust conceptual understanding without asking students to *reason with the concepts* in some way.

Understanding student reasoning on physics problems has long been a goal of physics education research. Early investigations of student conceptual understanding identified specific reasoning difficulties as well as conceptual difficulties and even found similar reasoning-related difficulties in different conceptual domains [23–26]. Observed difficulties were described and the empirical findings were used to guide the development of content-specific, research-based instructional materials [12,27,28].

Other early investigations sought to understand the composition of student conceptions of physics and to explain how or why certain conceptions were formed, cued, and used for reasoning [13,29–33]. These investigations created a framework that allows one to identify and observe the use of student “resources” for reasoning. It is posited by this framework that the act of reasoning is an act of cognitively selecting and coordinating, at the moment, the use of a subset of available resources.

Recently, there has been interest in investigating predictive control mechanisms that govern reasoning in a physics context [9,11,18,34]. Much of this work draws on findings from cognitive science and the psychology of reasoning. This strand of research has called for new methodologies to be employed in physics education research that would allow for the collection of data not normally accessible from a written response or think-aloud interview alone, but that would, to a greater degree, separate reasoning skills from conceptual understanding [9,10,16].

All of the work mentioned above has focused primarily on student responses to qualitative physics problems that require a series of inferences to be chained together into a line of reasoning. This ability to chain together a series of inferences is important while reasoning about physics concepts, and as such, studying it in more detail is necessary for progress in improving student performance in a physics classroom. This work aims to provide a new methodology for studying qualitative, inferential reasoning.

B. Other discipline-specific, reasoning-related research

The scope of reasoning-related research can be rightfully extended to other, expansive domains of research, such as student problem solving, scientific reasoning skills, mathematization, and others, but these domains have less relevance to the current work. Research on student problem solving and mathematization [35–39] emphasizes traditional quantitative problems that typically require the manipulation of multiple equations and quantities and seeks to understand and improve the strategies students employ while working through these problems. Likewise, there has been research related to scientific reasoning skills such as control of variables, conservation of volume, and proportional reasoning, and assessments have been used to study differences in proficiency with these skills between populations before and after instruction [23,40–43]. However, while quantitative problems and scientific reasoning are essential to a physics curriculum, the focus of this manuscript is on the structure of qualitative inferential reasoning patterns.

The *proof literature* in mathematics education research is more closely aligned with the specific goals of the investigation described in this article. Selden and Selden provide a review of this literature [44]. In a typical undergraduate mathematics program, there are specific courses that aim to teach students how to create mathematical proofs. These proofs tend to take the form of a series of deductive, qualitative inferences that are linked together as an argument in support of a specific conclusion. The research regarding student skill at constructing proofs is reminiscent of many research endeavors in physics education. Often, students' responses to a particular proof task are examined through various epistemological and conceptual lenses, with an emphasis placed on the identification of student difficulties with constructing proofs.

While the nature of the reasoning chains examined in the proofs literature is very closely related to those considered in this article, our work takes a different approach. Instead of examining possible causes for a particular reasoning difficulty, the current work aims to identify patterns in the structure of the reasoning chain itself; our goal is to provide new forms of data that can be utilized by future researchers investigating the mechanisms behind student construction of reasoning chains. As such, the current work would be very useful to researchers involved in studying student construction of mathematical proofs. Elements of a proof could be cast as a chaining task, and the resulting structures could be studied using the methodology described herein.

C. Network analysis in physics education research

In network analysis, objects of study are represented as nodes (dots), and a relationship between two given objects is represented as a link (a line connecting the two dots). A collection of nodes and links forms a network, and various parameters of the network can be analyzed.

Network analysis is fairly new to physics education research but has recently been seeing a dramatic increase in use, mostly in social network analysis characterizing the social dynamics within a physics community (i.e., a classroom, department, or university) and sometimes relating these dynamics to performance and learning gains within a physics course [45–51]. For instance, if studying the social dynamics of a classroom, the nodes could be the individual students in the classroom, with links formed for students who indicate that they have worked together in a meaningful way. However, network analysis has also been used to study epistemological shifts in conversations as a result of instruction (with the nodes being the topic of conversation and the links being formed when topics are discussed in close proximity) [52]; to model differentiation of concepts (where the nodes are concepts and the links are shared attributes of the two concepts) [53]; to assess patterns in representation used throughout a course employing modeling instruction (with the nodes being the type of representation used and links formed when two representations are used together) [54]; and to gain insight into the structure of answer patterns on a concept inventory (in one study, the nodes were students *and* responses to FCI questions, and the links were formed between the student and their responses) [55–57]. In each of these, the network nodes are different. In the literature regarding the resources framework, the coordination of resources has been studied using network-like representations, sometimes called "resource graphs" [58–61]. The current work utilizes network analysis to study the structure of student reasoning chains, which we believe is a novel pursuit; however, there are also connections to be made with resource graphs.

D. Summary and articulation of research questions

The data collection and analysis methodology presented in this manuscript is designed to separate, to the extent possible, reasoning skills from conceptual understanding and to provide data not normally accessible from written responses and think-aloud interviews. We aim to create a tool that can be used to study specific reasoning skills and to provide insight into the development of these skills. The main goal of this paper is to demonstrate how network analysis of reasoning chain construction tasks may be used to accomplish both objectives. As such, our investigation centered on the following research questions:

RQ1. To what extent can network analysis methodologies applied to chaining task data characterize the nature of student reasoning on qualitative physics questions?

RQ2. To what extent can our methodologies be used to track and document the development of a specific line of reasoning over the course of a two-semester introductory physics sequence?

III. METHODOLOGY

This section is broken into two main parts. In the first, we describe the reasoning chain construction task, which underlies the methodology employed here. In the second, we describe the network analysis methods that are used in this manuscript.

A. Reasoning chain construction tasks

A reasoning chain construction task, or *chaining task*, is a modified card-sorting task in which we (i) provide the student with a list of reasoning elements; (ii) indicate that all of the statements within these elements are true; and (iii) ask the student to construct a solution to a physics problem by selecting elements from the list, ordering them, and incorporating provided connecting words ("and," "so," "because," "but") as needed. The reasoning elements primarily consist of observations about the problem setup, statements of physical principles, and qualitative comparisons of quantities relevant to the problem, all of which are true. Everything the student needs to produce a complete chain of reasoning is present in the elements; the student's task is then to pick from the given conceptual pieces and directly assemble a reasoning chain. (The connecting words help people express the argument but are ultimately not used in the analysis presented in this manuscript.)

Reasoning chain construction tasks have primarily been implemented online using Qualtrics' "Pick/Group/Rank" question format [62]. This online format is illustrated in the context of a graph task and is shown in Fig. 1. Reasoning elements from the "Items" column, connecting words, and final conclusions can be dragged and dropped into the "Reasoning Space" box; the box increases in size vertically as elements are added.

These tasks were administered on participation-based homework assignments or exam reviews for students enrolled in the introductory calculus-based physics sequence, along with other questions relevant to the course but not relevant to the content found in the research tasks. The responses of all students completing a task in a given semester will be referred to as a single dataset.

Students received participation credit for completing these assignments in most cases, although extra credit (based on participation only) was awarded in some cases. In all cases, the tasks were administered after relevant lecture, laboratory, and small-group recitation instruction at a research-intensive university in New England. Research-based materials from *Tutorials in Introductory Physics* [6] were used in the course recitations. While course-specific demographic data could not be obtained, insight can be gained from institutional demographics during the years of the study. The institution had a population of undergraduate students where 53% identified as male and 47% identified as female; additionally, these students self-reported as White (83%), Hispanic/Latino (4%), multiracial (3%), Black/African American (2%), Asian (2%), and American Indian/Alaskan Native (1%).

The reasoning elements provided to the student were typically based on previously obtained student responses to open-ended, free-response versions of the task. Elements consisted of statements of first principles, observations about the task, and statements derived from first principles and observations. Some were productive to the correct line of reasoning, and some were not. Among the unproductive elements were elements that, while true, were useful primarily in constructing a common incorrect line of reasoning if there was one associated with the task. In addition, the extent to which students selected unproductive elements not associated with the correct or common incorrect line of reasoning could help us gauge the likelihood that students were simply inserting elements at random. Three blank elements labeled "Custom:" were provided, with instructions that students could use the text box attached to the custom element to create their own reasoning elements if they felt they wanted to add something not represented among the given reasoning elements. In practice, these elements ended up being used either for additional connecting words or for students to forgo the chaining aspect of the task altogether and instead type out a "paragraph style" response. These kinds of behaviors were typically observed in less than 5% of responses for any given task.

An important aspect of a chaining task is the intended logical connections between the provided reasoning elements—that is, the logical topology of the elements. For instance, some physics tasks require only a few steps to arrive at a correct answer (e.g., a qualitative question that can be solved via a short, linear chain of elements like the task shown in Fig. 1), while others require the student to combine two independent lines of reasoning (e.g., synthesis problems such as those reported by Ref. [63]); by casting each of these types of questions as a chaining task, we can obtain information about how students approach these different reasoning scenarios.

The provided reasoning elements determine to a large extent how students interact with the task. The elements were written by researchers (i.e., the authors of this article) who likely have a specific epistemological stance in mind, as well as a particular pedagogical perspective. The elements and especially the wording of the elements reflect the researchers' values about such ideas as what constitutes reasoning, a reasoning element, and the size of logical steps. For instance, an element corresponding to Newton's

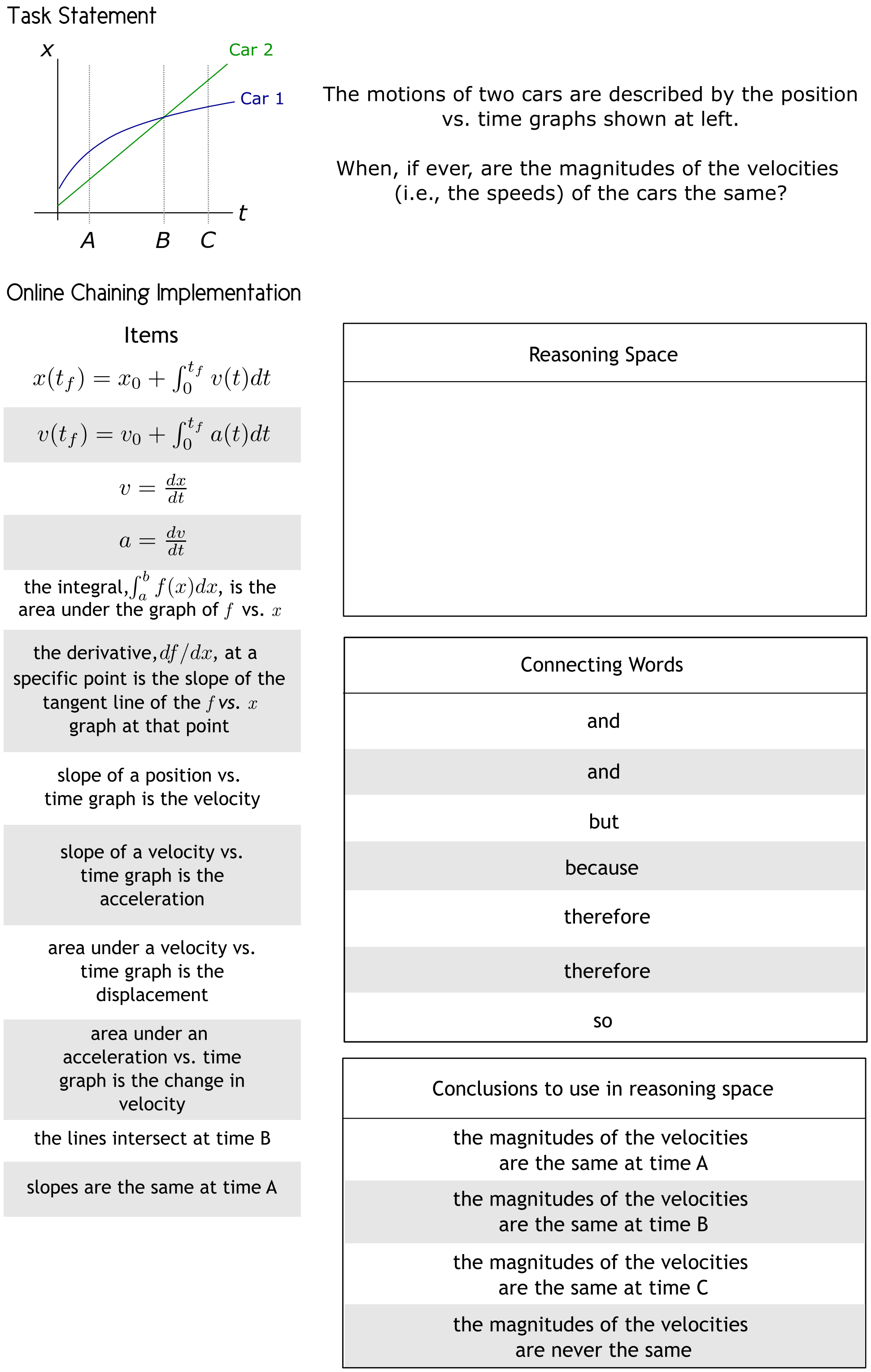


FIG. 1. An example of a reasoning chain construction task implemented online using Qualtrics' "Pick/Group/Rank" question format. This task is the same task discussed in Sec. IV B.

second law could read, among other things, "$F_{\rm net} = ma$," "the net force is equivalent to the mass times the acceleration," or "an acceleration is caused by a net force (distributed over a mass)." Each of these may convey a different meaning to the student, may interact differently with the context of the problem, and may differently represent what a "first principle" is and looks like. Thus, when interpreting responses to a chaining task, the main

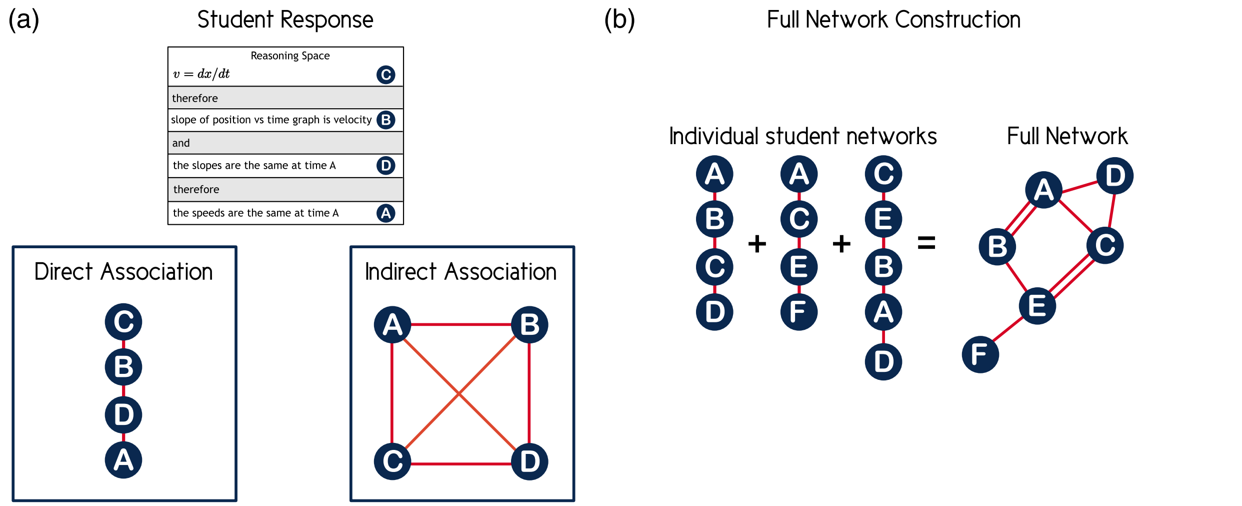


FIG. 2. (a) An example of two methods for constructing an individual-student network from an individual student's response. (b) An illustration of summing individual student networks to create a *full* network.

research endeavor is to ascertain not only how students reason generally about the problem but also how students engage in the specific types and lines of reasoning supported by the elements. In the tasks presented in this article, attempts were made to make the reasoning space topology as close to the observed student reasoning topology by drawing upon student written explanations of reasoning.

The chaining task (especially when implemented online) creates an environment in which students are required to present their argument in a linear progression of inferences, and this presentation of reasoning is separate from the process of reasoning that occurs in the mind. For instance, a student may consider a lengthy line of reasoning but feel that simplicity and elegance are valued in the sciences and therefore seek to construct the most concise argument possible with the elements; another student, though, may report a short chain out of a desire to get through the task quickly, without deep study of the elements provided. Regardless of these differences, there is still something valuable to be gained from analyzing patterns in the reasoning chains constructed by students. For example, suppose a student does not endorse first principles in their chain. We cannot assume that they did not consider first principles, but we can assume that if they did consider first principles, they made a decision (whether conscious or not) to exclude those considerations in the presentation of their reasoning.

1. Chaining task data as networks of associations

Chaining task data can be cast as a network for quantitative analysis. To accomplish this, the reasoning elements can be represented as nodes in a network, and associations made by the student between the elements can be represented as links, with the number of links between the nodes being called the *weight* of the link. We considered two main methods for establishing associations (links) between reasoning elements (nodes). These two methods are illustrated in Fig. 2(a). In the first, a connection is said to exist between two elements if the two elements are placed consecutively in a student's chain or on either side of a connecting word; a network created using this definition of association is referred to in this paper as a *direct association* network. In the second method, a connection exists between two elements if they appear together in the same student response; a network constructed in this way is referred to as an *indirect association* network. As we will describe below, both types of networks have different affordances, and exploring both is useful in interpreting student responses. In either method, individual student response networks are summed to create a network of all responses in a given dataset, as illustrated in Fig. 2(b). This is referred to as a *full* network in this manuscript.

In both methods, we remove connecting words from the data and use undirected links to form our networks. The connecting words, while serving in many cases to clarify the logic of a student's argument, posed a challenge for network analysis for two reasons. We observed that students often used connecting words intermittently and inconsistently. Additionally, even when connecting words were used, there remained ambiguity in the components that were intended to be associated with the connective, particularly when a task required multiple inferences. For instance, consider the phrase "A because B and C, therefore, D". This phrase could be parsed logically as "A because (B and C), therefore, D" or it could be parsed as "(A because B) and C, therefore, D". (Similar ambiguity exists regarding the parsing of the "therefore" connective prior to the D.) For these reasons, we felt uncomfortable attributing representational meaning to the connecting words when constructing the networks and decided to eliminate the connecting words from our analyses.

Because we removed the connecting words from students' responses when constructing a network, we also opted to make the links undirected. By choosing undirected links, we interpret a link between reasoning elements as simply a general "association" between those elements rather than interpreting any sort of logical meaning from the link. However, we find that this method of constructing networks does yield interpretable results, and we view this decision as a ground-level analysis of reasoning chains. Future analyses may be performed to investigate the usefulness of directed networks.

B. Network analysis

In this section, we present a brief overview of the network analysis techniques employed in this work. A detailed and more technical overview of each analysis technique is given in Appendix A. Later sections will describe in detail how to interpret the results of these methods in the context of reasoning chain construction tasks. All algorithms were implemented in *Wolfram Mathematica*.

1. Locally adaptive network sparsification (LANS)

Network sparsification aims to uncover the "backbone" structure of a large network by deleting links (sometimes called edges) that are unimportant to that structure [64]. In this study, we employed locally adaptive network sparsification (LANS) [64]. In LANS, the statistical significance of each link is calculated for the two nodes locally and a link is deleted only when it is found to be below a threshold value, α, of significance to both nodes. This preserves local structure in a network even if that local structure does not have as much weight as other parts of the network.

For the work presented here, the threshold α was chosen by lowering the threshold as much as possible before either nodes or collections of nodes began to be separated from the network. For instance, in some networks, there are elements that are more tightly associated with each other than with the rest of the network, and these may break off during sparsification when the threshold is too low. We wished to preserve the structure of the network to the extent possible while still simplifying it, so we felt uncomfortable breaking the network into separate pieces. Typical values of α for this work ranged from 0.1 to 0.2. These values ended up being consistent with those from other studies using LANS [64].

2. Community detection and bootstrap verification

The techniques of network analysis allow us to quantitatively determine groupings of elements, or *communities*, which are more tightly associated with each other than with the rest of the network. There are many methods of community detection available, and there is no single "best" method [65]. The method used in this work is called optimum modularity community detection [66]. This method of community detection was chosen based on its potential for interpretability of results and because the underlying statistical nature of the method allowed it to be useful for a broad range of network types. It was also selected because the method allowed for a rigorous definition of a community as an indivisible subgraph of the network (see details in Appendix A).

Before relying on the results of community detection, it is helpful to gauge how robust the community structure is. Could small perturbations produce a different community structure in the network? If the answer is yes, then it would be reasonable to question the divisions made by optimizing modularity. However, if the structures are impervious to random insertions or deletions, this would be clearer evidence of true community structure. To assess robustness, we employ a technique based on statistical bootstrapping that has been modified from Ref. [65] for the context of chaining tasks.

Our bootstrapping method involves creating a hypothetical dataset (with an N value equal to that of the original dataset) comprised of responses drawn at random from the actual student responses—but with a so-called catch and release approach, in which a individual student response may be included more than once in the hypothetical dataset. Then, a network is created from the hypothetical dataset and community detection employed. Typically, we generate 1000 hypothetical datasets. On each iteration, it is possible to test which elements are found in the same community. We considered an element to be part of a community if it is found in that community in at least 60% of the hypothetical datasets of the bootstrap test. By taking note of all the communities and their members in each iteration, a frequency plot can be generated showing how often a particular element is found in the same community as a test element. We therefore call these *reasoning element frequency plots.* An example of such a frequency plot is shown in Fig. 6.

3. Network measures: Centrality and clustering

Two network measures, *betweenness centrality* and *global clustering coefficients*, were utilized in the current work and will be described here. *Betweenness centrality* [67] is seen as a measure of a node's control over the "flow" in the network. A node's betweenness was originally defined as the number of shortest distance paths through that node divided by the total number of shortest distance paths in the network [67]. This definition applied only to unweighted networks, and so the definition was modified to respect the weights of the various links in the network [67], and we use a weighted betweenness centrality in this study.

The goal of a *global clustering coefficient* is to quantify how interconnected a network is. The clustering coefficient was originally defined as the number of closed triads (grouping of three nodes all connected to each other) divided by the total number of triads, either open (i.e., only two links among the three nodes) or closed (i.e., all nodes connected) [68]. The direct association network shown in Fig. 2(a) would have a clustering coefficient of 0, while the indirect association network shown in the same figure would have a clustering coefficient of 1. We use a clustering coefficient that has been extended to weighted networks [68]. Using a weighted coefficient, if a network had few closed triads but these triads weighted more heavily in the network, this network would rightly be considered interconnected.

IV. RESEARCH TASKS

In this section, we present network analyses of two different chaining tasks in physics in order to highlight the power of these methods in providing insight into student reasoning. The first task is set in a *work and energy* conceptual domain and introduces the interpretations of the network analysis methods in the context of chaining tasks. We then apply the network analysis methods to a set of four isomorphic graph-based tasks that span four content areas: kinematics, potential energy, electric potential, and magnetic flux. Network analysis of these graph-based tasks reveals the development of a more coherent line of reasoning across two semesters of introductory physics instruction.

A. Work-energy task

In order to highlight how the network analysis methods might be interpreted, we sought a task for which there was strong student performance as well as multiple, independent ways of answering correctly. This would de-emphasize conceptual difficulties and enable us to focus on the articulation of known correct knowledge. The goal of such a task would be to answer the following question: How effective are network analysis methodologies at characterizing and differentiating among different lines of reasoning on a physics question that most students can answer correctly? Analysis of this kind of task thus provides a good testing ground to determine the extent to which network analysis methodologies applied to chaining task data can characterize the nature of student reasoning on qualitative physics questions (RQ1).

Here we focus on a chaining task in the context of work and energy, and we use this task as an example of how the methods of network analysis can be interpreted in the context of chaining tasks. In this section, we describe the task, provide the results of the network analysis techniques described in Sec. III. B, and discuss the insights gained from this approach.

1. Physics question overview

The work-energy task was adapted from a concept question appearing in Knight's *Physics for Scientists and Engineers 4th ed.* [69]. In the task, students are told that a point particle moving to the left is slowing down because of a force pushing to the right, and no other forces are acting on the particle. Students are asked if the work done on the particle by the force is positive or negative or if there is not enough information to tell. The complete prompt as well as the reasoning elements provided to the student are shown in Fig. 3.

The correct answer is that the work on the particle by the force is negative. There are two viable ways of answering this question. The first involves recognizing that the work done is defined as the dot product between the force and displacement vectors and that a dot product of two vectors pointing in opposite directions is negative, thereby establishing that the work is similarly negative. This line of reasoning will be referred to as the *work as a dot product argument*. The second line of reasoning utilizes the principle of energy conservation and will be referred to as the *work as a change in energy argument*. This line uses a statement of the principle of energy conservation (i.e., $W_{\mathrm{net,external}} = \Delta U + \Delta K$) along with the observation that the particle is slowing down to argue that the work done on the particle by the force must be negative. For this approach, it is necessary to also specify that the kinetic energy is decreasing and that a point particle has no change in potential energy. This line of reasoning could be simplified by invoking the work-energy theorem (i.e., $W_{\mathrm{net}} = \Delta K$), which only applies to objects that may be treated as point particles, and omitting arguments related to potential energy.

Based on our analysis of student responses to similar questions in other formats, the most common incorrect response involves concluding that the work on the particle by the force is positive because the force is pushing to the right, which is assumed to be the positive direction. Similar findings have been reported by others [70].

2. Chaining task implementation

The reasoning elements provided to students on the chaining version of the work-energy task, shown in Fig. 3, were expressly designed to reflect both the *work as a dot product argument* and the *work as a change in energy argument*. While the common incorrect line of reasoning may also be constructed from the elements provided, all of the reasoning elements (with the exception of the incorrect conclusion elements) are true statements.

3. Performance overview

Of the 119 students who completed the chaining version of the work-energy task, 92% of them answered correctly that the work done by the force on the particle is negative. Of these correct responses, 69% responded with the *work as a dot product argument*, 12% responded with the *work as a change in energy argument*, and 16% included both

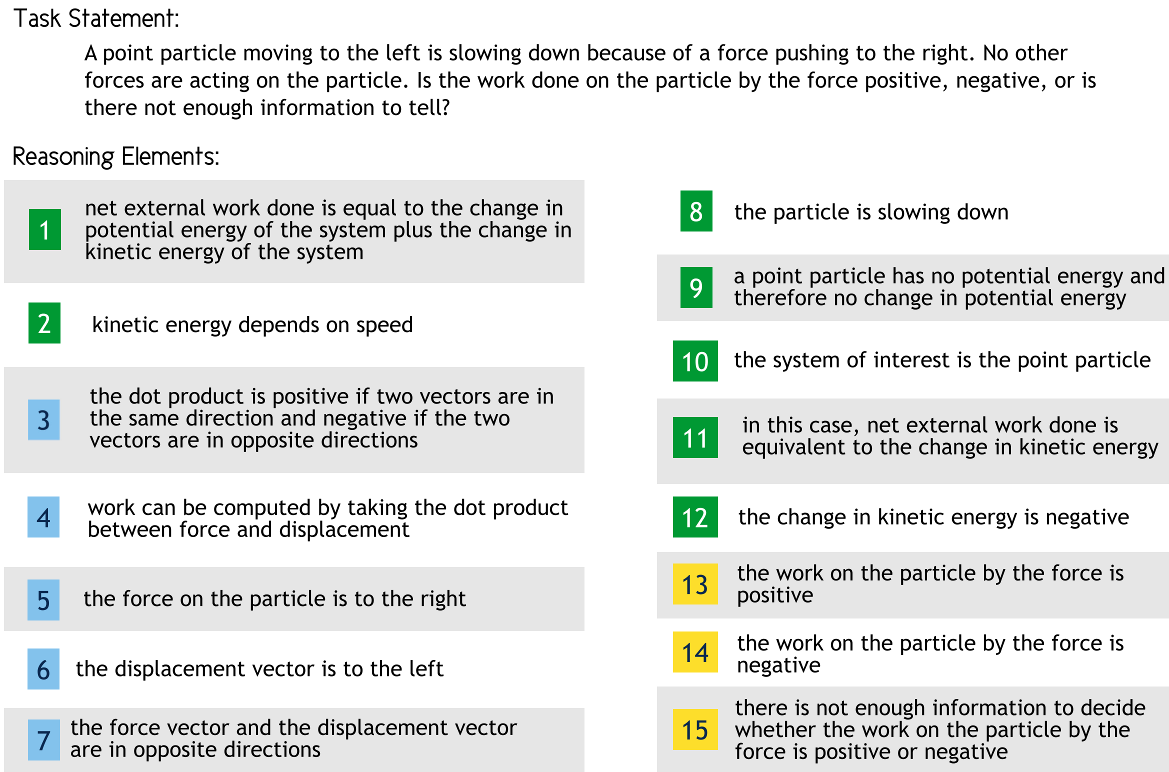


FIG. 3. Work-energy task. Question prompts (drawn from Ref. [69]) and associated reasoning elements provided to students are shown. The elements are numbered for later reference and color coded based on whether they were intended for the *work as a change in energy* argument (green), were intended for the *work as a dot-product* argument (blue), or were conclusion elements (yellow). Students were not presented with the colors associated with each element.

arguments. Figure 4 shows an example of each type of student response.

We have purposefully chosen to introduce network analysis using the work-energy task due to the unambiguous nature of the collected dataset, as this allows us to demonstrate the applicability and power of the network analysis tools before examining more complex, nuanced datasets. Because of the strong overall performance on the work-energy task, it is likely that students had a solid grasp of the reasoning involved in answering the question, and we therefore expected this to be reflected in their reasoning chains. Furthermore, since many students articulated each independent argument (energy and/or dot product), we recognized that these lines of reasoning would be clearly represented in a network constructed from all student responses. As a result, this set of student responses serves as an ideal test case for the application of the network analysis methods described above in the context of reasoning chain construction tasks.

4. Community detection analysis of correct responses

We constructed both a direct and an indirect association network from the correct responses to the work-energy task and applied the community detection algorithm to each separately. (Recall that, as discussed in Sec. III. A. 1, a direct association network only links elements that are placed consecutively in a student response, while an indirect association network links each reasoning element in a response with every other reasoning element in that response.) The results from that analysis are shown in Fig. 5. In the figure, the elements that are important to the *work as a dot product argument* are colored blue and the elements important to the *work as a change in energy argument* are colored green.

In both the direct and indirect association networks, the elements in the *work as a dot product argument* and the elements in the *work as a change in energy argument* are found to be separate from each other by the community detection algorithm. Additionally, the community structure of the direct association network reveals that the *work as dot product* elements appear to have two groupings: one with the two elements that state that the force vector is to the right and the displacement vector is to the left, and one with the rest of the *work as dot product* elements. In both networks, the answer element was found to be in the same community as the *work as a dot product* element. This is likely due to that argument being used more often among the responses.

We wish to note here that these results show that the two types of networks, direct and indirect, yield differing levels

FIG. 4. Examples of each type of response to the work-energy task. Each example is an actual student-generated chain.

of detail and indeed different types of information about the set of student responses represented. Thus, it is valuable to examine both types of networks. More will be said about this in Sec. IV A 6. We also wish to highlight some limitations of the specific visual representation shown here. In the network plots shown in Fig. 5, the weight of *within*-community edges is represented as multiple lines connecting two nodes, while *between*-community edges are collapsed to a single line. This means that information about the weight of those edges is obfuscated in this representation. However, we stress that the plots shown are a representation of the underlying community detection algorithm that determined that those weights were insignificant compared to the within-community weights that are represented. Thus, while the information missing from this representation may be of interest, it is not needed for the claims made in this paper.

(a) Bootstrapping community detection results. To assess the stability of the communities found via the optimum modularity community detection algorithm, bootstrap tests were administered by repeatedly testing hypothetical networks constructed from resampled correct responses, as explained in Sec. III B. We first discuss our examination of the communities arising in the direct association network and then turn our attention to the communities in the indirect association network.

For the direct association network, in every bootstrap test, the elements associated with the *work as a change in energy* argument and the *work as a dot-product* argument were well separated from each other. For example, consider

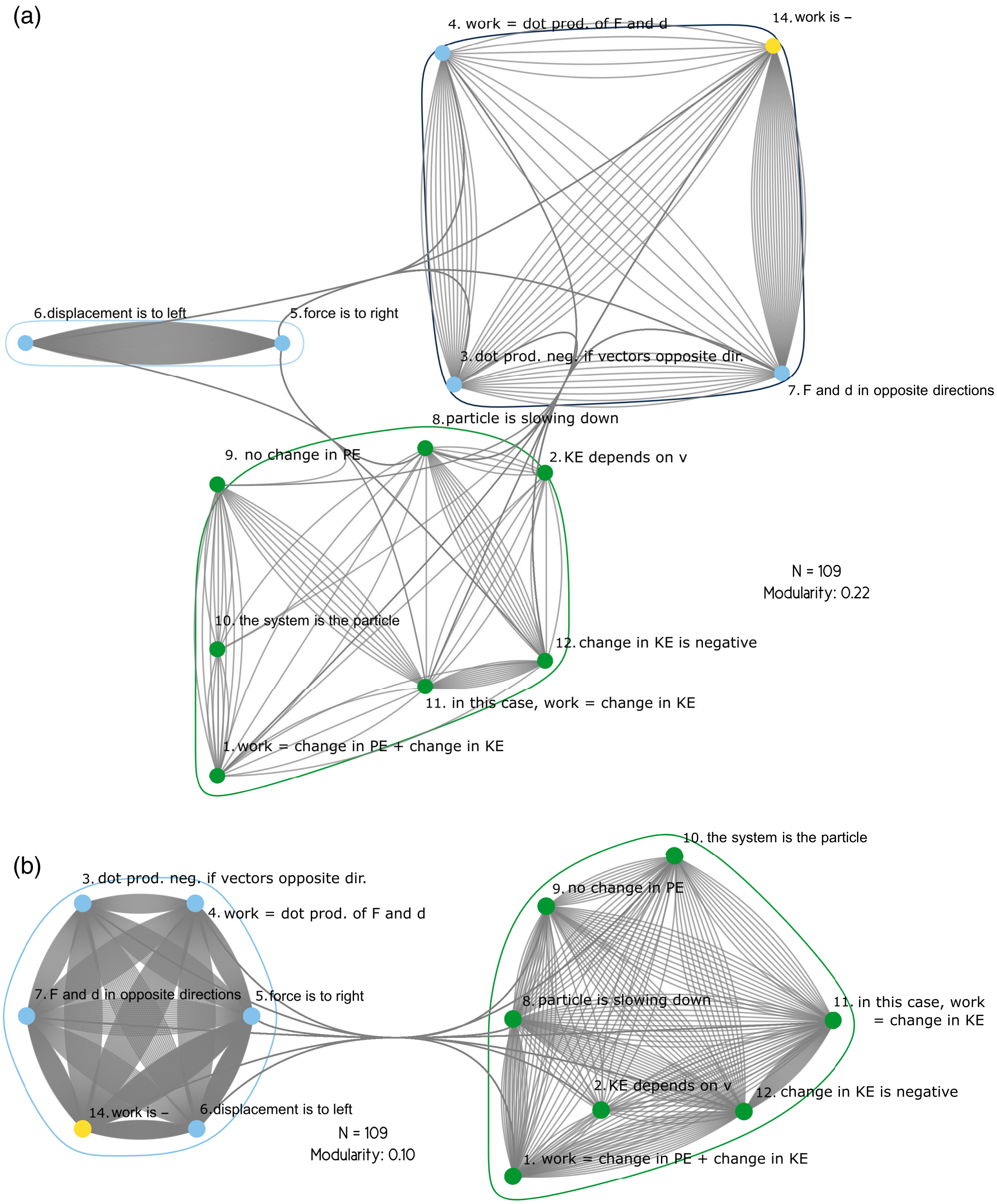


FIG. 5. A representation of the communities found in (a) a direct association network and (b) an indirect association network, both built from correct responses to the work-energy task described in Sec. IV. A. 2. Elements that are aligned with a *work as dot-product* argument are colored blue and the elements aligned with the *work as a change in energy* argument are colored green. The answer element is colored yellow. The modularity is given for reference. Note the especially dense connections between elements 5 and 6, suggesting a tight association between those elements in (a).

the reasoning element frequency plots shown in Figs. 6(a) and 6(b). The plots indicate the percentage of bootstrapping trials in which each element was included in a specified community. For these tests, an element that was found to be in the same community as the general statement of the principle of energy conservation (i.e., element 1) was considered to be a member of the *work as a change in energy* community. Similarly, elements found in the same community as the statement of work as a dot product (i.e., element 4) were considered members of the *work as a*

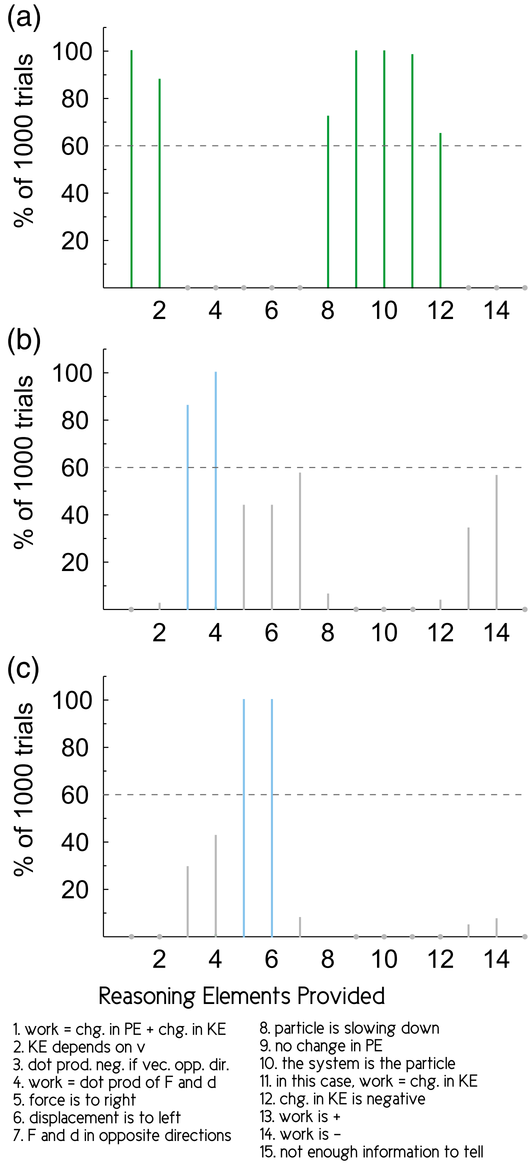


FIG. 6. Reasoning element frequency plot for three communities present in the direct association network, including (a) the *work as a change in energy* community, (b) the *work as a dot-product* community, and (c) the two-element force and displacement community [see Fig. 5(a)]. The plot indicates the percentage of the trials in which each element was included in the specified community. A dotted line corresponds to the 60% threshold used for ascertaining community membership in the bootstrapping tests.

dot-product community. The frequency plots reveal that the two arguments are well separated in the network since no element associated with the *work as a change in energy* argument appears in the *work as a dot-product community*, and vice versa, in close to 100% of the trials.

The reasoning element frequency plot for the two-element community shown in Fig. 5(a) [Fig. 6(c)] shows that the two elements "the force on the particle is to the right" (element 5) and "the displacement vector is to the left" (element 6) are always coupled together in the same community (1000 times out of 1000) but that between 30% and 40% of the time, the elements concerning the dot product (elements 3 and 4) are also included. These results indicate that this two-element structure is indeed present in the network.

It should be noted that in the reasoning element frequency plot for the *work as a dot-product* community [Fig. 6(b)], two elements (element 7 and element 14) approach, but remain below, the 60% threshold for membership in that community. Our threshold was set at 60% to ensure that elements were in the community more than half the time, but one could argue that these two elements also belong in the community or that the threshold should have been more restrictive—say 75% of iterations. The points made above about community membership in the *work as a change in energy* community are valid for a range of thresholds and do not rely solely on the 60% threshold set for this study. Additionally, the main point we are making here—that the bootstrap frequency test gives meaningful information about the two-element substructure shown in Fig. 5(a)—is valid also for a range of thresholds because the two elements in the substructure (elements 5 and 6) are well below the threshold for membership [around 40% as shown in Fig. 6(b)] and therefore are more easily seen as primarily belonging to an alternative community.

Based on the reasoning element frequency plots for the indirect association graph (see Supplemental Material [71]), all of the *work as a change in energy* argument elements are found 100% of the time in the community with the statement of the principle of energy conservation, and the elements related to the *work as a dot product* argument are likewise found 100% of the time with the statement of work as a dot product. Thus, we felt very confident in the robustness of the community structure depicted in Fig. 5(b).

5. Network sparsification method applied to work task correct responses

We now explore the usefulness of network sparsification by analyzing a direct association network built from the correct responses to the work task. Figure 7 shows a sparsified version of the direct association network at a threshold of $\alpha = 0.2$. The elements in this figure are color coded according to the same color scheme used in Fig. 3.

In Fig. 7, the two independent arguments are again separated as distinct in the network since the elements associated with the energy argument are separate from the elements associated with the dot-product argument (with the exception of element 8, which seems to serve as a bridge as will be discussed below).Furthermore, examination of the network reveals the existence of two clear chains of reasoning, each of which appears to include general principles (such as the principle of energy conservation or

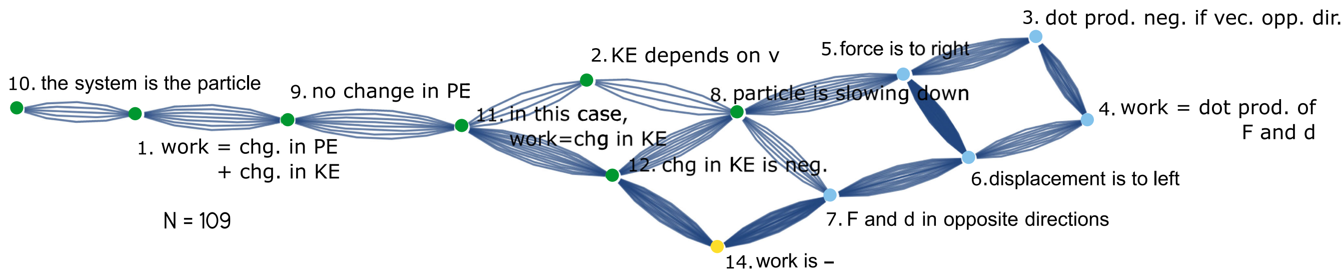


FIG. 7. A representation of a sparsified ($\alpha = 0.2$) direct association network built from correct responses to the work task. The elements are color coded according to the line of reasoning they are useful for: green elements are useful in the energy argument, and blue elements are useful in the dot-product argument.

the definition of work as the dot product of the force and displacement vectors) and to subsequently step through the application of the specifics in the problem statement before finally arriving at an answer. By qualitative inspection of the responses, it was seen that the element "the system of interest is the point particle" (element 10) was indeed a common starting point for students, as well as "the dot product is…" and "work can be computed by…" (elements 3 and 4). Additionally, the answer element was a common end point. Thus, based on the sparsified undirected graph and this additional information, the students in this case appeared to generally be starting with first principles and then applying situation-specific constraints to arrive at an answer.

(a) Assessing the fidelity of the sparsified representation. While the features of the sparsified graph are of interest, it is also good to assess, to the extent possible, whether they are true representations of the network structures or whether they are artifacts of the sparsification process. To assess the fidelity of the sparsified representation, we compare features of the sparsified network to network measures applied to the unsparsified network.

The first feature of interest is the observed topology of the network. The topology of the *work as a change in energy* argument elements, shown in Fig. 7, is observed to be quite linear, while the topology of the elements associated with the *work as a dot-product* argument is more interconnected. These apparent topological differences are reflected in the global clustering coefficients for each argument (recall from Sec. III. B. 3 that a global clustering coefficient quantifies how interconnected a network is). To determine if the two separate arguments (*energy* vs *dot product*) in the unsparsified network had the same qualitative level of interconnectedness, we compared *subnetworks* of the original unsparsified network. To create a subnetwork, we took the original unsparsified network of student responses to the work task and deleted all the nodes that did not pertain to a specific argument. We created two subnetworks for comparison: (i) a subnetwork of just the *work as a change in* energy elements, and (ii) a subnetwork of just the *work as a dot-product* element. Analysis of an unsparsified subnetwork composed of solely the elements in the *work as a change in energy* argument yields a clustering coefficient of 0.48. The global clustering coefficient of an unsparsified subnetwork consisting of just the elements in the *work as a dot-product argument* is 0.89—substantially higher. Thus, the relative interconnectedness of each of these arguments in the original, unsparsified networks (indicated by the clustering coefficients) appears to be preserved even after the sparsification process (indicated by the topology of the sparsified network); this consistency highlights both the fidelity and reliability of the chosen sparsification technique in retaining key characteristics of the network structure.

Another observed feature of the network structure is that element 8, "the particle is slowing down," bridges the two independent arguments. We sought to ascertain whether or not this element also served as a bridge in the unsparsified network. Bridges tend to have higher betweenness centrality as they are essential to the flow of information through a network (upon which the betweenness centrality is based), which means that betweenness centrality is a good measure to assess whether the feature is a bridge in the unsparsified network. The two elements in the unsparsified network with the highest betweenness are "the change in kinetic energy is negative" (element 12) and "the particle is slowing down" (element 8). These two elements, incidentally, have the same betweenness. Furthermore, in the sparsified network, those two elements also have the highest betweenness centrality. Thus, the unsparsified and sparsified networks share topological features that suggest to us that the sparsified structures are reliable representations of the original network structures on the basis of betweenness centrality as well.

The location of "the particle is slowing down" as a bridge in the network may be attributed to that particular element being used frequently in both the *work as a dot-product* argument and the *work as a change in energy* argument. Upon more detailed analysis of student responses, it was found that in the *work as a change in energy* argument, the element was used to justify why the kinetic energy (and thus the work) is negative, whereas in the *work as a dot-product* argument, the element was used to describe the consequence of the force and displacement being in opposite directions. This latter use may have stemmed, in part, from students referencing the task prompt, which noted that the particle "is slowing down because of a force pushing to the right."

6. Discussion of results

The separation of the elements into two distinct lines of reasoning in both the community detection results and the sparsification results shows that network analysis of chaining task data can explore the content of students' various arguments in a meaningful way. Additionally, the results show the role that each type of network (indirect vs direct association) can play in examining student reasoning. Based on our analyses, finding communities in the indirect association network seems best suited for determining which lines of reasoning are present among the responses, whereas community detection applied to direct association networks allows for greater resolution of the subarguments that make up those lines of reasoning.

Bootstrapping is an indispensable part of community detection. The reasoning element frequency plot revealed a stable subargument structure in the direct association network comprised of the elements "the force on the particle is to the right" (element 5) and "the displacement vector is to the left" (element 6). We would expect those two elements to be more closely associated with each other in the network since they were often placed next to each other in student responses, both in the chaining format and in free-response versions of this question. Indeed, the algorithm is sensitive to that structure.

The sparsified network appears to give information about how students viewed the structure of an argument. The linearity of the *work as a change in energy* argument and the nonlinearity of the *work as a dot-product* argument suggest a difference in how students approached those two arguments. The linearity or nonlinearity of the associations between a group of elements indicates that many students either responded with similar ordering of the elements (creating a linear network) or that there was not a preference for which elements came before others in the reasoning chain (creating a clustered, nonlinear network). It could be that this is inherent to the elements provided or it could be indicative of a particular learned approach to a problem.

Even if the specific interpretation of the structure is not always immediately apparent, the ability to quickly and efficiently characterize how a large group of students is approaching a line of reasoning can be very useful to instructors and researchers alike.

It is important to note, however, that the clear chain of reasoning shown in the sparsified graph does not necessarily represent the chain of reasoning constructed by the majority of individual students. Actually, only 2 students out of 100 responded with chains that included the first four elements of the energy argument (namely, elements 1, 9–11) in the order represented in Fig. 7, and only 8 used all four elements in their chain. Many students only cited parts of the argument, inserted irrelevant elements into their argument, arranged the argument differently, etc.; still, these students constructed their arguments in a way that led to the majority of the associations being between those four elements in the ordering shown in Fig. 7. Thus, the sparsified network represents a "wisdom of the crowd" result [72,73], a synergistic classroom consensus on how the elements ought to be arranged that transcends the reasoning chains constructed by individual students.

Further evidence of this synergistic consensus or wisdom of the crowd is provided by the results of the betweenness calculations. In the full, unsparsified network of correct student responses, the element "the particle is slowing down" served as a bridge between the two independent arguments and therefore has a high betweenness centrality. However, while that particular element was used by 27 students, only 2 students used the element in between the two arguments in their reasoning chain. Instead, the element's high betweenness centrality offers a glimpse into how the students as a whole viewed that particular element; in the logical landscape of this problem, the information that the speed is decreasing can be seen as relevant to both arguments. An implication of observing a dual-relevancy element is that the identified element may serve as a possible pivot point for shifting from one argument to the other during, for example, a classroom discussion of the solution to the task.

This classroom consensus reasoning can be useful in identifying where a class stands with respect to the usage of certain arguments. For instance, the work task was administered to two different calculus-based introductory mechanics courses at the same university, but with different instructors who had different instructional emphases. The sparsified network shown in Fig. 7 was derived from student responses during one of these courses and represents a full *work as a change in energy argument*, whereas the sparsified network of responses from the other class (see Supplemental Material [71]) gave a truncated *work as a change in energy* argument that only associates the elements "in this case, the net external work done is equivalent to the change in kinetic energy" (element 11) and "the change in kinetic energy is negative" (element 12) before arriving at an answer. The *work as a dot-product argument*, however, appeared to have been articulated in full by students in that same class. We speculate that the dissimilarities in the *work as a change in energy* arguments between the two courses are due to known differences in how each of the instructors approached problem solving with work and energy conservation. Such an effect has been noted in the literature [74]. However, we cannot rule out other factors such as the epistemological stance of the instructor and/or students, mastery of work-energy-related content, how each instructor graded reasoning for partial credit, etc. Our network data alone cannot isolate the reason for the difference, but they do provide a method of quickly ascertaining the nature of the difference. Thus, we find chaining tasks coupled with network analysis to be a useful diagnostic tool in investigating student reasoning patterns throughout instruction.

B. Isomorphic graph tasks

In this section, we report on student reasoning on a collection of four similar tasks administered over the course of two subsequent semesters of introductory calculus-based physics. Each of the tasks was designed to foreground the same line of reasoning in four different contexts. This experiment aimed to answer research question RQ2: "To what extent can our methodologies be used to track and document the development of a specific line of reasoning over the course of a two-semester introductory physics sequence?" Network analysis of these four tasks provided evidence that students developed a more sophisticated line of reasoning over the course of instruction. In this section, we focus specifically on the community detection technique in order to highlight this application of network analysis, but for brevity's sake, we do not discuss the other network measures.

1. Physics question overview

As part of an investigation of the impact of salient distracting features on patterns of student reasoning in the context of introductory physics, we developed four chaining-format graph tasks that are isomorphic in structure and are based upon one task in the literature, which we refer to as the kinematics graph task [9,18,32,75,76].

In the kinematics graph task, shown in Fig. 8, students are asked to determine when the speeds of two cars are the same by examining a plot of position vs time with two graphs representing the motion of the two cars. At time A, the slopes of the two graphs are the same, and at time B, the two graphs intersect. The correct answer is obtained by observing that the velocity is the derivative of position with respect to time, which on a graph corresponds to the slope of the tangent line at a point. Comparing slopes allows students to determine that the speeds (i.e., the magnitudes of the velocities) are the same at time A. As has been previously documented [9,18], however, many students answer that the speeds are the same at time B, the intersection point of the two graphs. The patterns of incorrect responses on these types of graphs have led to researchers investigating "slope-height confusion" and other difficulties related to interpreting and using graphs in a physics context [75–77], and these graph tasks have also been used to examine the impact of salient distracting features and domain-general reasoning phenomena on student performance in physics contexts [9,18].

All four tasks are structurally parallel, requiring students to recognize that a desired quantity can be obtained from

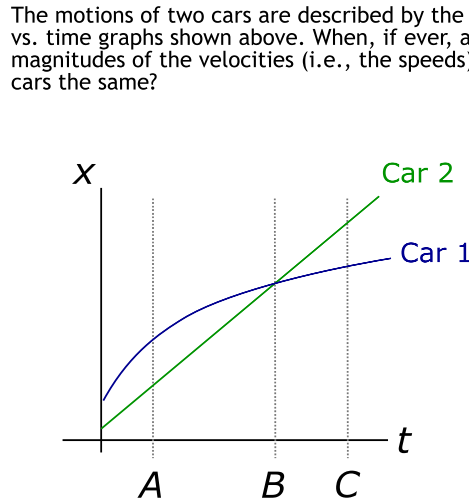


FIG. 8. The first of four isomorphic graph tasks adapted from Ref. [9]. The other three graph tasks are shown in detail in Appendix A. Note that this is the same task as that shown in Fig. 1.

the derivative, i.e., slope, of a given graph. The differences are in the given quantities being graphed on the x and y axes and the quantity that can be obtained from the slope of the graph. In addition to kinematics, graph tasks were constructed to highlight the relationship between force and potential energy, electric potential and field, and magnetic flux and electromotive force (emf). The tasks were presented to students near the end of instructional coverage of kinematics, potential energy, electric potential, and magnetic flux, respectively. (The other three tasks are presented, in detail, in Appendix B, Fig. 12.)

2. Chaining task implementation

The reasoning elements provided to the student in each task have been modified to fit the context but remain isomorphic in their structure. The reasoning elements are shown in Fig. 9. Unlike the work-energy task discussed in the previous section, these isomorphic tasks include a large number of elements that are irrelevant to both the correct and common incorrect lines of reasoning; indeed, 7 of the 12 elements are not relevant to any common line of reasoning. Note that many of these irrelevant elements are not truly isomorphic across tasks (for instance, the element "$a = dv/dt$" constitutes the second derivative of position, whereas the element "$F = dp/dt$" does not correspond to a second derivative of potential energy. However, all relevant elements, the surface features of the tasks, and the underlying structure of the correct line of reasoning are isomorphic.

There is an inherent logical structure among the productive elements provided to the students (shown in red in Fig. 9). While, at first glance, it may appear that the elements "$v = dx/dt$," "the derivative, df/dx, at a specific

	Kinematics Reasoning Elements	Potential Energy Reasoning Elements	Electric Potential Reasoning Elements	Magnetic Flux Reasoning Elements
1	$x(t_f) = x_0 + \int_0^{t_f} v(t)dt$	$U(x_f) = U_0 + \int_0^{x_f} \vec{F}(x) \cdot d\vec{x}$	$V(x_f) = V_0 + \int_0^{x_f} E(x)dx$	$\Phi_B = -\int_0^{t_f} \mathcal{E}dt$
2	$v(t_f) = v_0 + \int_0^{t_f} a(t)dt$	$p(t_f) = p_0 + \int_0^{t_f} F(t)dt$	$U(x_f) = U_0 + \int_0^{x_f} F(x)dx$	$\mathcal{E} = -\int_0^{s_f} \vec{E} \cdot d\vec{s}$
3	$v = \frac{dx}{dt}$	$F = -\frac{dU}{dx}$	$E = -\frac{dV}{dx}$	$\mathcal{E} = -\frac{d\Phi_B}{dt}$
4	$a = \frac{dv}{dt}$	$F = \frac{dp}{dt}$	$F = -\frac{dU}{dx}$	$E = -\frac{d\mathcal{E}}{ds}$
5	the integral, $\int_a^b f(x)dx$, is the area under the graph of f vs. x	the integral, $\int_a^b f(x)dx$, is the area under the graph of f vs. x	the integral, $\int_a^b f(x)dx$, is the area under the graph of f vs. x	the integral, $\int_a^b f(x)dx$, is the area under the graph of f vs. x
6	the derivative, df/dx, at a specific point is the slope of the tangent line of the f vs. x graph at that point	the derivative, df/dx, at a specific point is the slope of the tangent line of the f vs. x graph at that point	the derivative, df/dx, at a specific point is the slope of the tangent line of the f vs. x graph at that point	the derivative, df/dx, at a specific point is the slope of the tangent line of the f vs. x graph at that point
7	slope of a position vs. time graph is the velocity	the negative of the slope of a potential energy vs. position graph is the force	the slope of an electric *potential* vs. position graph gives the magnitude of the electric field	slope of a magnetic flux vs. time graph is the magnitude of the induced EMF
8	slope of a velocity vs. time graph is the acceleration	slope of a momentum vs. time graph is the force	the slope of an electric *potential energy* vs. position graph is the force	slope of an induced EMF vs. position graph is the magnitude of the induced electric field
9	area under a velocity vs. time graph is the displacement	area under a force vs. position graph is the work done by the force, which is the negative of the change in potential energy	area under an electric field vs. position graph is the electric potential	area under an induced EMF vs. time graph is the magnitude of the change in magnetic flux
10	area under an acceleration vs. time graph is the change in velocity	area under a force vs. time graph is the change in momentum (or the impulse)	area under a force vs. position graph is the electric potential energy	area under an induced electric field vs. position graph is the magnitude of the change in the induced EMF
11	the lines intersect at time B	the lines intersect at position B	the lines intersect at position B	the lines intersect at time B
12	slopes are the same at time A	slopes are the same at position A	slopes are the same at position A	slopes are the same at time A
13	the speeds are the same at time A	the magnitudes of the forces are the same at position A	the magnitudes of the electric fields are the same at position A	the magnitudes of the induced EMF's are the same at time A
14	the speeds are the same at time B	the magnitudes of the forces are the same at position B	the magnitudes of the electric fields are the same at position B	the magnitudes of the induced EMF's are the same at time B
15	the speeds are the same at time C	the magnitudes of the forces are the same at position C	the magnitudes of the electric fields are the same at position C	the magnitudes of the induced EMF's are the same at time C
16	the speeds are never the same	the magnitudes of the forces are never the same	the magnitudes of the electric fields are never the same	the magnitudes of the induced EMF's are never the same

FIG. 9. Reasoning elements provided to the student on each of the four isomorphic graph tasks. Elements that are productive to a correct line of reasoning are color coded in red, elements that are not productive are color coded in blue, and answer elements are color coded in yellow.

point is the slope of the tangent line of the f vs x graph at that point", and "slope of a position vs time graph is the velocity" are equivalent and interchangeable statements, they actually constitute a logical argument justifying why the slope is the velocity; namely, the two elements "$v = dx/dt$" and "the derivative[…] is the slope…" combine to imply the third element. We refer to the collection of these three elements as the *velocity triad* (even outside of the kinematics context). We also refer to the element "slope […] is the velocity" as a *derived heuristic* because it represents a chunked knowledge piece [78] that is derived from two independent principles. While it would be acceptable to many instructors if students were to simply use the *slope as a velocity*-derived heuristic, all three elements are needed to provide a logically sound argument. Their inclusion, then, provided an opportunity for additional insight into whether students tend to justify their arguments with first principles or instead rely on derived heuristics practiced in class.

3. Performance overview

Given the contexts associated with these isomorphic tasks, data were collected in both semesters (fall and

TABLE I. Overview of student performance on the four isomorphic graph tasks. The correct answer choice is indicated in boldface.

Response	Kinematics ($N = 121$)	Potential energy ($N = 183$)	Electric potential ($N = 77$)	Magnetic flux ($N = 187$)
Time A (correct)	**63%**	**58%**	**62%**	**78%**
Time B (intersection)	33%	36%	31%	19%
Time C	0%	3%	4%	3%
Never	4%	3%	3%	0%

spring) of the on-sequence calculus-based introductory physics course. All tasks were administered after relevant course instruction. Chronologically, the kinematics task was administered first in the year, the potential energy task second, the electric potential task early in the second semester of physics, and the magnetic flux task last. Because the four graph tasks were administered across a single academic year, most students who completed the introductory calculus-based sequence would have seen and completed multiple, and likely all four, tasks.

Student performance for these tasks is shown in Table I. The percentage of responses answering correctly increased very slightly over the two-course sequence, but the salient distracting feature (the intersection point) remained a strong distractor, with approximately 20% or more of students answering consistently while attending to the intersection point.

4. Arguments identified via community detection

Each indirect association network (see Supplemental Material [71]) built from all responses (correct and incorrect) to the graph tasks generally breaks into two communities: the correct answer community, which always includes the elements isomorphic to "$v = dx/dt$," "slope of a position vs time graph is the velocity," and "slopes are the same at time A" and sometimes includes the "derivative is slope" element; the common incorrect answer community, which includes the element isomorphic to "the lines intersect at time B"; and all of the other elements in a loosely connected network. Interestingly, the element "the derivative, df/dx, at a specific point is the slope of the tangent line of the f vs x graph at that point." (element 6), which is very relevant to the correct line of reasoning, was not found in the correct answer community for the kinematics and potential energy graph task but was found in the correct answer community in the electric potential and magnetic flux task. We would have expected this element to always be associated with the correct answer.

To investigate this phenomenon more fully, we examined community structure in indirect association networks comprised of *just the correct responses* to each task. The resulting networks are shown in Fig. 10. The elements that make up the full, detailed correct line of reasoning are colored red in the figure, while all other elements are colored dark blue except the answer element, which is colored yellow. The *derivative is slope* element is not in the main answer community for the first two graph tasks but becomes more tightly associated with the correct answer in the final two graph tasks. One may notice that, as mentioned in Sec. IV. A. 4, the particular representation, shown in Fig. 10, obfuscates the number of links between the *derivative is slope* element and the other red elements. As we did there, we note here that the underlying community detection algorithm determined that the links between the *derivative is slope* element and the other red elements were less significant than the links within the correct answer community circled in red.

A reasoning element frequency plot for the correct community, shown in Table II, revealed that the *derivative is slope* element is indeed increasing in use across the four tasks with the exception of the potential energy task and thus increasing over the course of the two-semester introductory calculus-based physics sequence. Again, as in Sec. IV. A. 4, this claim holds even if the 60% threshold for membership of the *derivative is slope* element in the correct answer community was either more restrictive (such as 75%) or less restrictive (50%).

The community structures of the direct association networks for the four graph tasks (see Supplemental Material [71]) also reveal a shift in how the *derivative is slope* element is used by students. In the responses to the kinematics and potential energy tasks, the element is not a member of the correct answer community or in the same community as the other productive elements, whereas in the responses to the electric potential and magnetic flux tasks, the element is more closely associated with the productive elements. The *derivative is slope* element appears to be an important indicator of the development of this line of reasoning. A particularly compelling community structure is found in the direct association network built from correct responses to the magnetic flux task, which asks students to identify the time at which two emfs are the same from a graph of magnetic flux as a function of time. This community structure (shown in the Supplemental Material [71]) shows a subcommunity made up of the "velocity triad" elements, which include the *derivative is slope* element. In a direct association network, a connection is only formed between two elements that are placed consecutively. Thus, the subcommunity of the "velocity triad" elements means that those three elements were consistently placed next to each other in student responses. This confirms the elevation of the *derivative is slope* element in the magnetic flux task compared to the kinematics and potential energy task.

We note that other interesting insights could be drawn from the sparsified networks of data from the graph tasks,

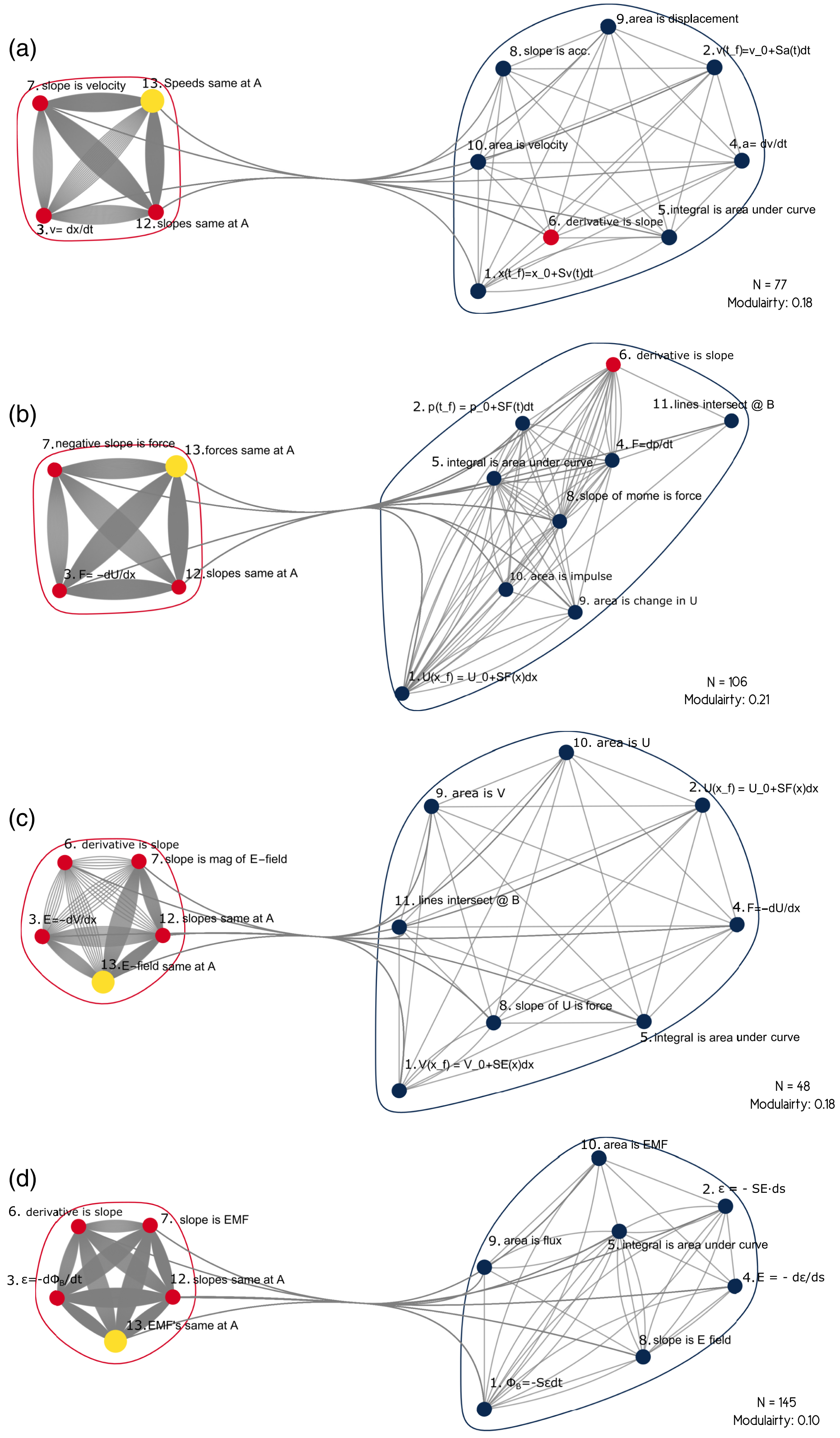


FIG. 10. Community structure detected in indirect association networks comprised of correct responses to the graph task as posed in the context of (a) kinematics, (b) potential energy, (c) electric potential, and (d) magnetic flux.

TABLE II. The results of bootstrapping test for the correct answer community. Results are shown in table form rather than a plot for ease of reading. Elements referencing velocity are in quotes as a reminder that in the nonkinematics graph tasks, this element contained words appropriate to the specific context.

	Kinematics	Potential energy	Electric potential	Magnetic flux
Derivative is slope	46%	29%	74%	100%
"$v = dx/dt$"	85%	100%	95%	100%
Slope is "velocity"	100%	100%	100%	100%
Slopes same at A	100%	100%	100%	100%

but our focus in this section has been on highlighting the affordances of community detection in a specific application.

5. Discussion of results

The results of network analysis of the four isomorphic graph tasks again demonstrate that community detection can meaningfully separate lines of reasoning in the responses according to the answer choice. Thus, the key result that network analysis of chaining task data provides useful and interpretable information is replicated in this task.

Perhaps the most important result from the isomorphic graph tasks is the observed development of a cohesive line of reasoning regarding the "velocity triad" of elements, seen in the community detection analyses presented in the previous section. The identified communities in both the direct and indirect association networks indicate that the *derivative is slope* element was not tightly associated with the other productive elements (including the correct answer element) for the mechanics tasks but was tightly associated with those elements for the electromagnetics tasks.

The proportion of correct responses that include the *derivative is slope* element is 14% for the kinematics task, 24% for the potential energy task, 24% for the electric potential task, and 27% for the magnetic flux task, indicating that the frequency of use overall is not increasing much over the last three tasks. Instead, the element must have been more frequently placed in correct responses that include only the other productive elements rather than being placed in responses that include unproductive elements as well—that is, the element is being used "more productively." Additionally, students must have been placing the *derivative is slope* element in closer proximity to the other productive elements since the direct association networks also saw a tighter coordination between those elements.

We propose that, as the sequence progresses, the students responding to these tasks either better understand the connection between that element and the other elements or are more comfortable with the use of that element alongside the other elements.

Why would this shift occur? One explanation for the relative nonuse of the element among correct respondents on the kinematics graph task is that the phrase "slope of a position vs time graph is the velocity" is often a "chunked" cognitive element or heuristic, even among experts.[1] We presume that the students who answer correctly on this task in the context of kinematics employ the learned heuristic that the slope of a position vs time graph is the velocity and ignore the first principles from which that heuristic is derived. When asked the question in a context in which they have not formed such a heuristic, they may then resort to a wider examination of the separate elements.

The heuristic may have been formed to varying degrees in other contexts. For instance, in the magnetic flux task, it may be that students were less familiar with the application of Faraday's law to a graph of magnetic flux than they were with, say, how to get an electric field from a graph of electric potential. Because of a lack of familiarity, students may have relied more on calculus to make a connection between Faraday's law and the graph, as opposed to simply knowing from the features of the graph how to obtain an answer. This is supported by a brief review of the curriculum. In the course textbook [69], there are many examples of switching between field and potential graphically, but most examples concerning Faraday's law were centered on nongraphical considerations. Thus, the heuristic was probably more familiar in the electric potential task than it was in the magnetic flux task, with both being less familiar than the kinematics task.

Another possibility is that the students, over the course of the two semesters, became more comfortable and/or more proficient with the language and concepts of calculus, such that they felt comfortable endorsing elements that explicitly included those concepts. As part of our validation efforts related to these reasoning chain construction tasks, some think-aloud interviews with students were conducted. These interviews seemed to support this interpretation as well, at least in the aspect of students not feeling comfortable with the language of calculus employed on the kinematics task. Further work would need to be done to determine the extent to which comfort with calculus impacts the use of the *derivative is slope* element. A significant percentage of students were concurrently taking the first calculus course as a corequisite at the time the kinematics task was administered, and derivatives were covered later in the semester in calculus relative to kinematics in physics.

[1]We have administered the chaining version of the kinematics graph task to physics and other STEM educators and a frequent comment we hear is that the three elements "$v = dx/dt$", "derivative is slope", and "velocity is slope" are functionally equivalent. Only when it is pointed out that the former two are independent statements that combine to justify the latter is it agreed upon that the three elements are actually logically different.

Evidence for a shift in the structural usage of the *derivative is slope* element across the four tasks appears across multiple metrics, underscoring the utility of the network analysis of chaining tasks for examining student formation of specific reasoning chains. While the cause of the shift cannot be ascertained from these data alone, the shift is evident in the communities detected in indirect association networks, as well as the communities detected in direct association networks, the sparsified versions of direct association networks, and the betweenness calculations on those direct association networks (the latter two are not discussed in this paper for brevity). Thus, network analysis techniques are sensitive to shifts in reasoning chains over time and, as such, could be used to gauge how students are building reasoning skills over time.

V. CONCLUSIONS AND FUTURE WORK

The overarching goal of this manuscript was to illustrate how a new methodology, network analysis of student responses to *reasoning chain construction tasks,* can generate valuable knowledge surrounding how students reason on physics questions, specifically those questions that require stepping through a series of qualitative inferences. As we have shown, network analysis of responses to chaining tasks generates novel data related to both the content and structure of student arguments. Here, we will discuss general affordances seen across both tasks and then highlight how these affordances, and other patterns observed in the data, can be used to bolster existing analysis methods or generate entirely new research questions.

Across all tasks, we have demonstrated that network analysis of chaining task data can separate lines of reasoning associated with different answers. Via community detection, we were able to find elements that were more tightly associated with a given answer than the other elements in the set (as in the case of the kinematics graph task); these tight associations were interpretable as typical reasoning seen from students in free response or interview settings. One affordance of the network methodology is that the categorization of the elements into lines of reasoning associated with a particular answer is automatic using the community detection algorithm, so large datasets can be analyzed quickly. Furthermore, by studying the community structure in both direct and indirect association networks, one can determine a set of elements that are core to an argument, and which are associated but somewhat peripheral to arriving at a particular answer. Clear distinctions between correct and incorrect arguments were also seen in the sparsification results of the graph tasks, indicating once again that the lines of reasoning associated with particular answers can be meaningfully separated in chaining task data.

Network sparsification yields further insight into another aspect of student reasoning with the provided elements: on each task shown, sparsification was meaningfully interpreted as the "wisdom of the crowd" consensus about the structure (or logical landscape) of the identified arguments. In most of the tasks on which we report, the structure of the associations among the elements revealed information that would not have been available from an examination of the responses individually. For instance, in the work-energy task, the linear structure of the *work as a change in energy* argument compared to the clustered structure of the *work as a dot-product* argument would have been hard to ascertain by simply studying the individual responses alone.

A further, perhaps more powerful, use of network analysis of chaining task data is to observe specific lines of reasoning before, during, and after instruction. The isomorphic graph tasks revealed that over the course of two semesters, a specific reasoning element representing a specific connection between calculus concepts and features of a graph (the *derivative is slope* element) was more productively incorporated into a line of reasoning relating velocity to features of a position graph. These results suggest that network analysis of reasoning chain construction task data can be used to isolate and study the development of specific reasoning skills. This could be helpful in assessing the impact of instructional materials on student reasoning with specific arguments. For instance, many instructional materials (especially scaffolded tutorials) step students through qualitative inferential arguments while developing physics conceptual knowledge or teaching problem-solving strategies. These same qualitative inferential arguments are then expected to be used on new but similar questions such as those found on exams. Chaining tasks could be used to study student use of these arguments before, during, and after instruction. Coupled with network analysis techniques, chaining tasks can be utilized to study many types of arguments, specifically arguments related to the reasoning difficulties identified in the physics education research literature.

Reasoning chain construction tasks may have the potential to investigate theories of student reasoning [18], especially when used in conjunction with network analysis. As one example, studies using the resources framework posit that different reasoning outcomes may share a subset of similar resources, with only one or two resources not in common with each other [59,61]. By studying patterns of element selection and, more particularly, how students arrange those elements, evidence for (or against) this view may emerge from reasoning chain construction tasks. The focus on the structure of associations made by students provides access to unique data regarding student qualitative reasoning and lends itself to the exploration of various theories of qualitative reasoning.

Network analysis of reasoning chain construction tasks has the potential to become a valuable tool for researchers in physics education. Here, we have demonstrated its affordances for facilitating the investigation of specific reasoning chains through novel data generation and informing instruction in new ways. Perhaps most importantly, it has the potential to be a distinct asset to ongoing efforts to investigate and strengthen student reasoning in physics, including those that attend to domain-general reasoning phenomena.

Finally, reasoning chain construction tasks represent a novel activity type for students that highlights inferential reasoning and could potentially be used as a formative tool for *teaching* students to attend to deductive processes. Chaining tasks are fairly simple to create. An instructor can readily review students' written responses to a specific question and construct a set of chaining task elements from those responses. While the network analysis techniques employed here are generally not feasible without access to and fluency with computational software, in the future, it is possible that more sophisticated online homework systems may have the ability to provide the results of the network analysis to interested instructors without the need for the instructor to actually perform that analysis. We have begun working to ensure that there will be a platform on which instructors can develop and administer online reasoning chain construction assessments with relative ease. Implementing chaining tasks in an online homework platform would also create the possibility of large datasets for which network analysis is particularly helpful in exploring. This has the potential to increase the generalizability of the results and reveal patterns across many different but related tasks.

ACKNOWLEDGMENTS

The authors wish to acknowledge valuable discussions with Eric Brewe and Zachary Hutchinson. The authors also appreciate substantive feedback about this work from John R. Thompson, Michael C. Wittmann, and Natasha M. Speer, and past and present members of the Physics Education Research Laboratory at the University of Maine. In addition, the lead author would like to acknowledge the University of Maine for support via a Signature and Emerging Areas of Excellence Graduate Fellowship in the area of STEM Education. This material is based upon work supported by the National Science Foundation under Grants No. DUE-1431940, No. DUE-1431541, No. DUE-1431857, No. DUE-1432052, No. DUE-1432765, No. DUE-2142416, No. DUE-2142276, No. DUE-2141975, No. DUE-2142436, and No. DRL-0962805.

APPENDIX A: NETWORK ANALYSIS TECHNICAL OVERVIEW

In this section, we present a technical overview of the network analysis techniques employed in this work.

1. Locally adaptive network sparsification (LANS)

Network sparsification aims to uncover the "backbone" structure of a large network by deleting links (sometimes called edges) that are unimportant to that structure [64]. One simple method for achieving this is to establish a threshold value for a link's weight and delete all links that fall below this threshold. For instance, one might decide a connection is only relevant if more than 5% of students made the connection, and so one would delete any link that had a weight less than the value of $0.05 \cdot N$, where N represents the population size. However, this method does not preserve some structures that are important to the network. Perhaps, a small group of students decided to be detailed in their reasoning chains, and so they added structure to the network that is relevant to overall patterns of reasoning but, due to its low prevalence among the whole population, this structure might get cut from the network by an arbitrarily set threshold weight. Additionally, it may be hard to guess, *a priori*, a threshold weight that preserves these structures and still reduces the complexity of the network.

A more sophisticated, method of sparsification is locally adaptive network sparsification (LANS) [64]. In LANS, the statistical significance of each link is calculated for the two nodes locally and a link is deleted only when it is found to be below a threshold value of significance to both nodes. This preserves the local structure that would be dismantled using a threshold link weight method. The LANS method is implemented by first calculating the fractional link weight of a link connecting nodes i and j, as

$$p_{ij} = \frac{w_{ij}}{\sum_{k=1}^{N_i} w_{ik}},$$

where w_{ij} is the weight of the link, and the sum in the denominator is over all the nearest neighbors of the node i. Then, the cumulative distribution function (CDF) is computed as

$$F_{ij} = \frac{1}{N_i} \sum_{k=1}^{N_i} \hat{1}\{p_{ij} < p_{ik}\},$$

and the link is retained if $F_{ij} > \alpha$, where α is the predetermined significance threshold. These same calculations are completed for every link in the network. The function $\hat{1}\{p_{ij} < p_{ik}\}$ returns the value 1 if the argument is true and 0 if the argument is false.

To give an example of how this method works, a sample network [Fig. 11(a)] was constructed, and the technique was applied. The main structure of the original network is represented by the lettered nodes. The link between nodes D and E is 7 times weaker than the link between nodes D and C; all other links between lettered nodes are roughly equivalent in strength. The added nodes 6–8 were given

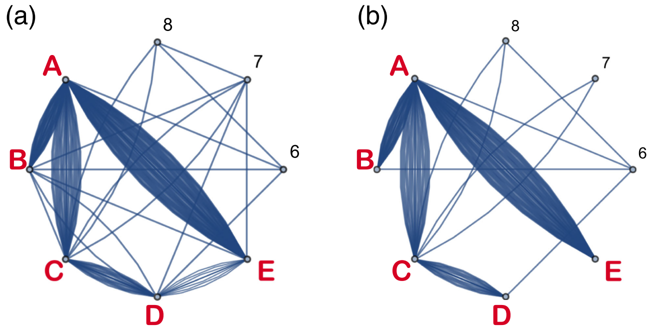


FIG. 11. Example network illustrating locally adaptive network sparsification [64]. (a) The base network. (b) The same network after sparsification at $\alpha = 0.1$.

random connections to each other and the other nodes in the network to simulate smaller structures that may be of interest and generate "noise." The sparsified network is shown in Fig. 11(b). The smaller structures have been retained even after the network has been simplified via the LANS technique but the connection between nodes D and E has been severed along with the weaker connections to node B (except the one to node 6). Thus, this technique is able to preserve small structures while still detecting and removing weaker connections among the larger structures.

Note that the four connections to node 6 remain. This is because those four connections are equally significant to node 6. More generally, anytime a node has only links of weight one, all of those links will be preserved due to the nature of the algorithm. Because of the tendency to automatically preserve nodes such as node 6, we "prune" sparsified networks by removing all links of weight 1 *after* sparsification to make the network more readable.

For the work presented here, the threshold α was chosen by lowering the threshold as much as possible before either nodes or collections of nodes began to be separated from the network. For instance, in some networks, there are elements that are more tightly associated with each other than with the rest of the network, and these may break off during sparsification when the threshold is too low. We wished to preserve the structure of the network to the extent possible while still simplifying it, so we felt uncomfortable breaking the network into separate pieces. Typical values of α for this work ranged from 0.1 to 0.2. These values ended up being consistent with those from other studies using LANS [64].

2. Community detection

The techniques of network analysis allow us to quantitatively determine groupings of elements, or *communities*, which are more tightly associated with each other than with the rest of the network. There are many methods of community detection available, and there is no single "best" method [65]. The method used in this work is termed optimum modularity community detection [66]. This method of community detection was chosen based on its potential for interpretability of results and because the underlying statistical nature of the method allowed it to be useful for a broad range of network types. It was also selected because the method allowed for a rigorous definition of a community as an indivisible subgraph of the network.

Network modularity is proportional to the number of links between a pre-defined group of elements minus the number of expected links in an equivalent network (i.e., one with the same nodes) in which the links are placed at random. The expected number of links is $k_i k_j/2m$, where k_i and k_j are the degrees of node i and node j, and m is the total number of links in the network and is given by $m = \frac{1}{2}\sum_i k_i$. Thus, the expected number of links is related to the degree of the node: the higher the degree, the more likely it is to have links in a network in which the links are placed at random.

The modularity is maximized by dividing the network into two subgraphs of maximum modularity and then repeating this process for each of the two parts. If any proposed division causes the total modularity to decrease, the corresponding subgraph is preserved and considered a community, and the algorithm moves on to the next subgraph until all communities are found. Thus, a community is defined as an indivisible subgraph of the network.

Before relying on the results of community detection, it is helpful to gauge how robust the community structure is. Could small perturbations produce a different community structure in the network? If the answer is yes, then it would be reasonable to question the divisions made by optimizing modularity. However, if the structures are impervious to random insertions or deletions, this would be clearer evidence of true community structure. To assess robustness, we employ a technique based on statistical bootstrapping that has been modified from Ref. [65] for the context of chaining tasks.

For a dataset of N student responses, our bootstrapping technique consists of creating a hypothetical dataset comprised of $M = N$ responses drawn at random from the N actual student responses. (A specific response in the original dataset may be selected more than once for the hypothetical dataset; if this were not the case, the hypothetical dataset would be equivalent to the actual dataset.) This hypothetical dataset is treated as a new dataset and a network is constructed from it. The community structure of this new hypothetical network is found, and tests are applied to the hypothetical community structure. The process is then repeated for many iterations, tallying the results of the tests to determine how frequent a particular result is. It is suggested to perform as many iterations as possible, but in chaining task analysis, the frequency of a particular result converges on a specific value in under 1000 iterations. Accordingly, in the research described in this manuscript, a standard 1000 iterations were found to be sufficient to obtain reliable information.

In this manuscript, we use the bootstrapping technique to test how often particular elements are found in the same community as a test element. We select an element of interest (such as an answer element, or one indicative of a specific argument being used) and determine which of the other elements are consistently in the same community as that element. By taking note of the community members in each iteration, a frequency plot can be generated from the results. We therefore call these *reasoning element frequency plots*. An example of such a frequency plot is shown in Fig. 6. We consider an element to be part of a community if it is found in that community in at least 60% of the iterations of the bootstrap test.

3. Network measures: Centrality and clustering

Two network measures, *betweenness centrality* and *global clustering coefficients*, were utilized in the current work and will be described here. *Betweenness centrality* [67] is seen as a measure of a node's control over the "flow" in the network. A node's betweenness was originally defined as the number of shortest distance paths through that node divided by the total number of shortest distance paths in the network [67]. This definition applied only to unweighted networks, and so the definition was modified to respect the weights of the various links in the network by defining "shortest distance" as a combination of the traditional "distance" (i.e., number of nodes on a path between two end nodes) and a "conductance" (i.e., the weighting of the different links on a path between two end nodes) [79]. The modification of betweenness of Ref. [67] for weighted networks relies on a similar definition of shortest distance and is represented as

$$d(i,j) = \min\left(\frac{1}{(w_{ih})^{\beta}} + \cdots + \frac{1}{(w_{hj})^{\beta}}\right),$$

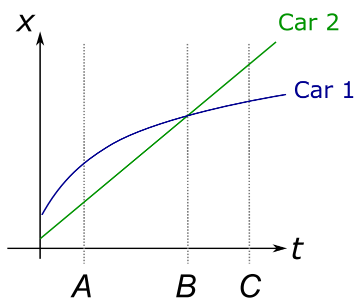
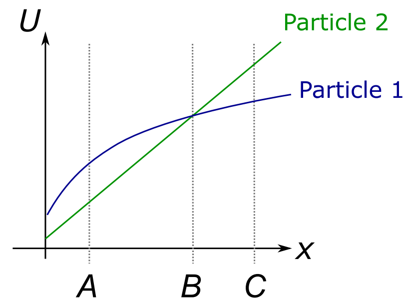
where d is the shortest distance between node i and node j, w_{gh} is the weight of the link between nodes g and h, and β is a positive tuning parameter set based on the context that the network represents. When $\beta < 1$, the number of nodes in a path becomes a greater influence on the distance, whereas for $\beta > 1$, the weight of the links becomes a greater influence. In chaining networks, the weight of a link represents the number of students who made an association between the two elements and so it should have the most influence over the distance: a path that many students established should be of smaller distance than a short path that only a few students took. However, we do not wish to completely drown out structures created by only a few students. For this reason, we selected a value of 1.5 for β. The betweenness is then calculated in the same manner as for unweighted graphs—by finding the ratio of the number of shortest paths through a given node to the number of shortest paths in the network.

Global clustering coefficients were also defined originally for unweighted networks and needed to be extended for weighted networks. The goal of a global clustering coefficient is to quantify how interconnected a network is. The clustering coefficient was originally defined as the number of closed triads (grouping of three nodes all connected to each other) divided by the total number of triads, either open, (i.e., only two links among the three nodes) or closed (i.e., all nodes connected) [68]. The direct association network shown in Fig. 1 would have a clustering coefficient of zero, while the indirect association network shown in that figure would have a clustering coefficient of 1. The idea of clustering is extended to weighted networks by assigning a weight, ω, to each triad in the network based on the weights of the links in the triad [68]. The weights, ω, are computed from the geometric mean of the weights of the two links stemming from the center node of the triad. The clustering coefficient can then be defined as follows, with τ representing the set of triplets and τ_{Δ} representing the set of closed triplets:

$$C_{\omega} = \frac{\text{total value of closed triplets}}{\text{total value of triplets}} = \frac{\sum_{\tau_{\Delta}} \omega}{\sum_{\tau} \omega}.$$

Thus, if a network had many closed triads compared to open triads, but the open triads were weighted more heavily, the network would not be considered interconnected. Conversely, if a network had few closed triads but these triads weighted more heavily in the network, this network would rightly be considered interconnected.

APPENDIX B: ISOMORPHIC GRAPH TASKS

Task	Kinematics Graph Task	Potential Energy Graph Task
Figure		
Task Statement	The motions of two cars are described by the position vs. time graphs shown above. When, if ever, are the magnitudes of the velocities (i.e., the speeds) of the cars the same?	The potential energy of system 1, in which only particle 1 can move, is described by the potential energy vs. position graph shown. Likewise, the potential energy of system 2, in which only particle 2 can move, is shown. The two systems don't interact. Where, if anywhere, are the magnitudes of the forces on the particles the same?

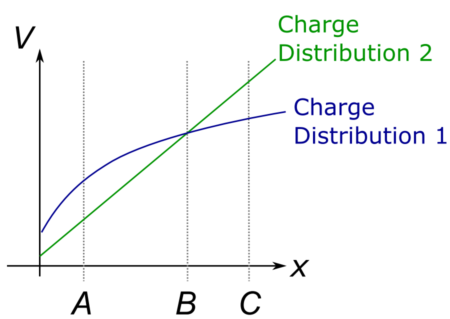
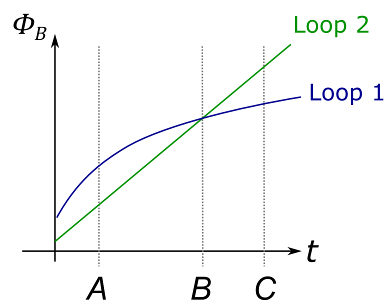
Task	Electric Potential Graph Task	Magnetic Flux Task
Figure		
Task Statement	The electric potentials set up by two charge distributions located far away from each other are described by the electric potential vs. position graphs shown above. Where, if anywhere, are the magnitudes of the electric fields due to the charge distributions the same?	The magnetic fluxes through two different conducting loops in different magnetic fields are described by the magnetic flux vs. time graphs shown above. When, if ever, are the absolute values of the induced EMF's ($\mathcal{E}_1$ and $\mathcal{E}_2$) the same?

FIG. 12. Four isomorphic graph tasks used in the study.

[1] N. D. Finkelstein and S. J. Pollock, Replicating and understanding successful innovations: Implementing tutorials in introductory physics, Phys. Rev. ST Phys. Educ. Res. **1**, 010101 (2005).
[2] J. M. Saul and E. F. Redish, Final evaluation report for FIPSE Grant #P116P50026: Evaluation of the Workshop Physics Dissemination Project (1997), https://physics.ucf.edu/~saul/articles/WP-FIPSE_Rprt.pdf.
[3] D. R. Sokoloff and R. K. Thornton, Using interactive lecture demonstrations to create an active learning environment, Phys. Teach. **35**, 340 (1997).
[4] R. Beichner, The Student-Centered Activities for Large Enrollment Undergraduate Programs (SCALE-UP) Project, in *Research-Based Reform of University Physics* (PER Central, 2007), Vol. 1.
[5] C. H. Crouch and E. Mazur, Peer Instruction: Ten years of experience and results, Am. J. Phys. **69**, 970 (2001).
[6] L. C. McDermott and P. S. Shaffer, *Tutorials in Introductory Physics* (Pearson College Division, Upper Saddle River, NJ, 2001).
[7] L. C. McDermott, *Physics by Inquiry: An Introduction to Physics and the Physical Sciences* (John Wiley & Sons, Inc., New York, 1995), Vol. 2.
[8] M. C. Wittmann, R. N. Steinberg, and E. F. Redish, Activity-Based Tutorials: Introductory Physics, *The Physics Suite* (Wiley, New York, 2004).
[9] A. F. Heckler, The role of automatic, bottom-up processes: In the ubiquitous patterns of incorrect answers to science questions, in *Psychology of Learning and Motivation* (Academic Press, Cambridge, MA, 2011), pp. 227–267.
[10] M. Kryjevskaia, M. R. Stetzer, and N. Grosz, Answer first: Applying the heuristic-analytic theory of reasoning to examine student intuitive thinking in the context of physics, Phys. Rev. ST Phys. Educ. Res. **10**, 020109 (2014).
[11] P. R. Heron, Testing alternative explanations for common responses to conceptual questions: An example in the context of center of mass, Phys. Rev. Phys. Educ. Res. **13**, 010131 (2017).
[12] L. C. McDermott, Oersted Medal Lecture 2001: Physics Education Research—The Key to Student Learning, Am. J. Phys. **69**, 1127 (2001).
[13] E. F. Redish, A theoretical framework for physics education research: Modeling student thinking, in *Proceedings of the International School of Physics Enrico Fermi* (2004), Vol. 156, pp. 1–63, https://files.eric.ed.gov/fulltext/ED493138.pdf.
[14] D. Kahneman, *Thinking, Fast and Slow* (Farrar, Straus and Giroux, New York, NYC, 2013).
[15] R. Rosiek and M. Sajka, Eyetracking in research on physics education, in *Key Competences in Physics Teaching and Learning*, edited by T. Greczyło and E. Dębowska, Springer Proceedings in Physics (Springer, Cham, 2016), pp. 67–77.
[16] J. R. Sattizahn, D. J. Lyons, C. Kontra, S. M. Fischer, and S. L. Beilock, In physics education, perception matters, Mind Brain Educ. **9**, 164 (2015).
[17] A. F. Heckler and A. M. Bodgan, Reasoning with alternative explanations in physics: The cognitive accessibility rule, Phys. Rev. Phys. Educ. Res. **14**, 010120 (2018).
[18] J. C. Speirs, M. R. Stetzer, B. A. Lindsey, and M. Kryjevskaia, Exploring and supporting student reasoning by leveraging dual-process theories of reasoning and decision making, Phys. Rev. Phys. Educ. Res. **17**, 020137 (2021).
[19] M. Kryjevskaia, M. R. Stetzer, B. A. Lindsey, A. McInerny, P. R. L. Heron, and A. Boudreaux, Designing research-based instructional materials that leverage dual-process theories of reasoning: Insights from testing one specific, theory-driven intervention, Phys. Rev. Phys. Educ. Res. **16**, 020140 (2020).
[20] C. Gette and M. Kryjevskaia, Establishing a relationship between student cognitive reflection skills and performance on physics questions that elicit strong intuitive responses, Phys. Rev. Phys. Educ. Res. **15**, 010118 (2019).
[21] J. C. Speirs, W. N. Ferm, M. R. Stetzer, and B. A. Lindsey, Probing student ability to construct reasoning chains: A new methodology, presented at *PER Conf. 2016, Sacramento, CA*, 10.1119/perc.2016.pr.077.
[22] M. L. Nagel and B. A. Lindsey, Implementation of reasoning chain construction tasks to support student explanations in general chemistry, J. Chem. Educ. **99**, 839 (2022).
[23] R. A. Lawson and L. C. McDermott, Student understanding of the work-energy and impulse-momentum theorems, Am. J. Phys. **55**, 811 (1987).
[24] M. E. Loverude, C. H. Kautz, and P. R. Heron, Helping students develop and understanding of Archimedes' principle. I. Research on student understanding, Am. J. Phys. **71**, 1178 (2003).
[25] C. H. Kauts, P. S. Shaffer, and L. C. McDermott, Student understanding of the ideal gas law, Part II: A microscopic perspective, Am. J. Phys. **73**, 1064 (2005).
[26] B. A. Lindsey, P. R. Heron, and P. S. Shaffer, Student ability to apply the concepts of work and energy to extended systems, Am. J. Phys. **77**, 999 (2009).
[27] L. C. McDermott, Millikan Lecture 1990: What we teach and what is learned—Closing the gap, Am. J. Phys. **59**, 301 (1991).
[28] P. R. Heron, Empirical investigations of learning and teaching, part I: Examining and interpreting student thinking, in *Proceedings of the International School of Physics Enrico Fermi* (2004), Vol. 156, pp. 341–350, 10.3254/978-1-61499-012-3-341.
[29] A. A. diSessa, Toward an epistemology of physics, Cognit. Instr. **10**, 105 (1993).
[30] A. A. diSessa and B. L. Sherin, What changes in conceptual change?, Int. J. Sci. Educ. **20**, 1155 (1998).
[31] D. Hammer, More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for education research, Am. J. Phys. **64**, 1316 (1996).
[32] A. Elby, What students learning of representations tells us about constructivism, J. Math. Behav. **19**, 481 (2000).
[33] D. Hammer, A. Elby, R. Scherr, and E. F. Redish, Resources, Framing, and Transfer, *in Transfer of Learning: Research and Perspectives* (Information Age Publishing Inc., Greenwich, CT, 2005), pp. 89–119.
[34] A. K. Wood, R. K. Galloway, and J. Hardy, Can dual-processing theory explain physics students' performance

on the Force Concept Inventory?, Phys. Rev. Phys. Educ. Res. **12**, 023101 (2016).

[35] L. Hsu, E. Brewe, T. M. Foster, and K. A. Harper, Resource letter RPS-1: Research in problem solving, Am. J. Phys. **72**, 1147 (2004).

[36] J. Tuminaro and F. R. Edward, Elements of a cognitive model of physics problem solving: Epistemic games, Phys. Rev. ST Phys. Educ. Res. **3**, 020101 (2007).

[37] T. J. Bing and E. F. Redish, Analyzing problem solving using math in physics: Epistemological framing via warrants, Phys. Rev. ST Phys. Educ. Res. **5**, 020108 (2009).

[38] J. L. Docktor, J. Dornfeld, E. Frodermann, K. Heller, L. Hsu, K. A. Jackson, A. Mason, Q. X. Ryan, and J. Yang, Assessing student written problem solutions: A problem-solving rubric with application to introductory physics, Phys. Rev. Phys. Educ. Res. **12**, 010130 (2016).

[39] E. M. Smith, J. P. Zwolak, and C. A. Manogue, Isolating approaches: How middle-division physics students coordinate forms and representations in complex algebra, Phys. Rev. Phys. Educ. Res. **15**, 010138 (2019).

[40] V. P. Coletta and J. A. Phillips, Addressing barriers to conceptual understanding in ie physics classes, in *Proceedings of the 2009 Physics Education Research Conference, Ann Arbor, MI* (AIP, New York, 2009).

[41] L. Bao, T. Cai, K. Koenig, K. Fang, J. Han, J. Wang, Q. Liu, L. Ding, L. Cui, Y. Luo, Y. Wang, L. Li, and N. Wu, Learning and scientific reasoning, Science **323**, 586 (2009).

[42] L. Ding, Verification of causal influences of reasoning skills and epistemology on physics conceptual learning, Phys. Rev. ST Phys. Educ. Res. **10**, 023101 (2014).

[43] S. White Brahmia, A. Olsho, T. I. Smith, A. Boudreaux, P. Eaton, and C. Zimmerman, The Physics Inventory of Quantitative Literacy: A tool for assessing mathematical reasoning in introductory physics, Phys. Rev. Phys. Educ. Res. (to be published).

[44] A. Selden and J. Selden, Overcoming students' difficulties in learning to understand and construct proofs, *Making the connection: Research and Teaching in Undergraduate Mathematics* (Mathematical Association of America, Washington, DC, 2008).

[45] E. Brewe, 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).

[46] J. Bruun and E. Brewe, 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).

[47] T. M. Sault, H. G. Close, and S. F. Wolf, Student cognition in physics group exams, presented at PER Conf. (2018), 10.1119/perc.2018.pr.Sault.

[48] D. L. Vargas, A. M. Bridgeman, D. R. Schmidt, P. B. Kohl, B. R. Wilcox, and L. D. Carr, Correlation between student collaboration network centrality and academic performance, Phys. Rev. Phys. Educ. Res. **14**, 020112 (2018).

[49] K. Commeford, E. Brewe, and A. Traxler, Characterizing active learning environments in physics using network analysis and classroom observations, Phys. Rev. Phys. Educ. Res. **17**, 020136 (2021).

[50] K. Commeford, E. Brewe, and A. Traxler, Characterizing active learning environments in physics using network analysis and classroom observations, Phys. Rev. Phys. Educ. Res. **17**, 020136 (2021).

[51] J. Pulgar, D. Ramirez, A. Umanzor, C. Candia, and I. Sanchez, Long-term collaboration with strong friendship ties improves academic peformance in remote and hybrid teaching modalitities in high school physics, Phys. Rev. Phys. Educ. Res. **18**, 010146 (2022).

[52] M. Bodin, Mapping university students' epistemic framing of computational physics using network analysis, Phys. Rev. ST Phys. Educ. Res. **8**, 010115 (2012).

[53] I. T. Koponen, Systemic view of learning scientific concepts: A description in terms of directed graph model, Complexity **19**, 27 (2013).

[54] D. McPadden, Examining students' representation choices in university modeling instruction (2018).

[55] E. Brewe, J. Bruun, and I. G. Bearden, Using module analysis for multiple choice responses: A new method applied to Force Concept Inventory data, Phys. Rev. Phys. Educ. Res. **12**, 020131 (2016).

[56] J. Yang, J. Wells, R. Henderson, E. Christman, G. Stewart, and J. Stewart, Extending modified module analysis to include correct responses: Analysis of the Force Concept Inventory, Phys. Rev. Phys, Educ. Res. **16**, 010124 (2020).

[57] J. Wells, H. Sadaghiani, B. P. Schermerhorn, S. Pollock, and G. Passante, Deeper look at question categories, concepts, and context covered: Modified module analysis of quantum mechanics concept assessment, Phys. Rev. Phys, Educ. Res. **17**, 020113 (2021).

[58] M. C. Wittmann, Using resource graphs to represent conceptual change, Phys. Rev. ST Phys. Educ. Res. **2**, 020105 (2006).

[59] T. I. Smith and M. C. Wittmann, Applying a resources framework to analysis of the force and motion conceptual evaluation, Phys. Rev. ST Phys. Educ. Res. **4**, 020101 (2008).

[60] M. S. Sabella and E. F. Redish, Knowledge organization and activation in physics problem solving, Am. J. Phys. **75**, 1017 (2007).

[61] K. Black and M. C. Wittmann, Procedural resource creation in intermediate mechanics, in *Proceedings of the 2009 Physics Education Research Conference, Ann Arbor, MI* (AIP, New York, 2009).

[62] Qualtrics, Provo, UT, USA (2019).

[63] B. Ibrahim, L. Ding, A. F. Heckler, D. R. White, and R. Badeau, Students' conceptual performance on synthesis physics problems with varying mathematical complexity, Phys. Rev. Phys. Educ. Res. **13**, 010133 (2017).

[64] N. J. Foti, J. M. Hughes, and D. N. Rockmore, nonparametric sparsification of complex multiscale networks, PLoS One **6**, e16431 (2011).

[65] S. Fortunato, Community detection in graphs, Phys. Rep. **486**, 75 (2010).

[66] M. E. J. Newman, Modularity and community structure in networks, Proc. Natl. Acad. Sci. U.S.A. **103**, 8577 (2006).

[67] T. Opsahl, F. Agneessens, and J. Skvoretz, Node centrality in weighted networks: Generalizing degree and shortest paths, Soc. Netw., **32**, 245 (2010).

[68] T. Opsahl and P. Panzarasa, Clustering in weighted networks, Soc. Netw. **31**, 155 (2009).

[69] R. D. Knight, *Physics for Scientists and Engineers: A Strategic Approach with Modern Physics*, 4th ed. (Pearson, 2016), Chap. 9.

[70] M. E. Loverude, C. H. Kautz, and P. R. L. Heron, Student understanding of the first law of thermodynamics: Relating work to the adiabatic compression of an ideal gas, Am. J. Phys. **70**, 137 (2002).

[71] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.20.010147 for additional network diagrams and frequency plots.

[72] F. Galton, Vox Populi, Nature (London) **75**, 450 (1907).

[73] J. Surowiecki, The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations, doubleday (2004).

[74] J. P. Zwolak and C. A. Manogue, Assessing student reasoning in upper-divison electricity and magnetism at Oregon State University, Phys. Rev. Phys. Educ. Res. **11**, 020125 (2015).

[75] L. C. McDermott, R. M. L., and E. H. van Zee, Student difficulties in connecting graphs and physics: Examples from kinematics, Am. J. Phys. **55**, 503 (1987).

[76] R. J. Beichner, Testing student interpretation of kinematics graphs, Am. J. Phys. **62**, 750 (1994).

[77] W. M. Christensen and J. R. Thompson, Investigating graphical representations of slope and derivative without a physics context, Phys. Rev. ST Phys. Educ. Res. **8**, 023101 (2012).

[78] J. D. Bransford, *How People Learn: Brain, Mind, Experience, and School* (National Academies Press, Washington, DC, 2000).

[79] M. E. Newman, The structure of scientific collaboration networks, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).