Network analysis of students' conceptual understanding of mathematical expressions for probability in upper-division quantum mechanics

William D. Riihiluoma,[1] Zeynep Topdemir,[2] and John R. Thompson[3,4]

[1]Department of Physics and Astronomy, Dickinson College, Carlisle, Pennsylvania 17013, USA
[2]School of Education, Johannes Kepler University Linz, Linz, Austria
[3]Department of Physics and Astronomy, University of Maine, Orono, Maine 04469, USA
[4]Maine Center for Research in STEM Education, University of Maine, Orono, Maine 04469, USA

(Received 21 February 2024; accepted 1 July 2024; published 15 July 2024)

One expected outcome of physics instruction is for students to be capable of relating physical concepts to multiple mathematical representations. In quantum mechanics (QM), students are asked to work across multiple symbolic notations, including some they have not previously encountered. To investigate student understanding of the relationships between expressions used in these various notations, a survey was developed and distributed to students at six different institutions. All of the courses studied were structured as "spins-first," in which the course begins with spin-$1/2$ systems and Dirac notation before transitioning to include continuous systems and wave function notation. Network analysis techniques such as community detection methods were used to investigate conceptual connections between commonly used expressions in upper-division QM courses. Our findings suggest that, for spins-first students, Dirac bras and kets share a stronger identity with vectorlike concepts than are associated with quantum state or wave function concepts. This work represents a novel way of using well-developed network analysis techniques and suggests such techniques could be used for other purposes as well.

DOI: 10.1103/PhysRevPhysEducRes.20.020102

I. INTRODUCTION

Physicists use mathematics for far more than computation: mathematical expressions and relationships are utilized to help them understand and reason about the world [1]. This is certainly the case in quantum mechanics (QM), in which, due to its often nonintuitive nature, one needs to rely on mathematical reasoning to understand and make predictions for systems on the quantum scale. The level of abstraction and mathematical sophistication used in upper-division QM coursework has been shown to present many challenges to students, including when interpreting Dirac formalism [2], reasoning about possible wave functions both symbolically and graphically [3], distinguishing between Euclidean and Hilbert spaces [3,4], and studying time dependence and time evolution [3–6]. Student understanding of representations of eigenequations has been studied by education researchers both in mathematics [7–9] and physics [10–12], as has the number of different notations that are frequently used and the varied mathematics that each notation requires [13–15]. These notations typically include Dirac, vector-matrix, and wave function notations—all of which require varied mathematical operations and understanding for fruitful application to QM systems.

An interesting aspect of studying representational understanding in upper-division QM is the variance in how and when the various aforementioned notations are introduced. Commonly used texts often begin by either introducing the Schrödinger equation and wave functions [16] or the Stern-Gerlach experiments (spin-$1/2$ systems) and Dirac notation [17,18]. These two different approaches are often referred to as "wave functions-first" and "spins-first," respectively, and though each text does eventually introduce the other notation, this can lead to courses with markedly different structures and notational emphases.

One challenge when studying student understanding at the upper division, in general, is the smaller sample size when compared to introductory physics courses. Due to attrition and a transition away from a general education audience, the number of students taking upper-division courses is naturally far smaller [19]; this typically manifests in research studies as a focus on more qualitative methodologies (clinical interviews being a classic example). While these methodologies are excellent opportunities for providing a deep view into individual students' conceptual understanding, they often also lead to a loss in the generalizability of claims that can be made. This loss in generalizability applies both within a given course (unless every enrolled student is studied) and across equivalent

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

courses at different institutions. One way to improve generalizability is to expand the data pool to include students at multiple institutions.

A technique that has seen increased use within the physics education research (PER) community in recent years is that of network analysis. Network analysis encompasses any technique that focuses on connections between different actors. Historically, these techniques were developed to study transportation and information networks [20] but have since been used within the PER community to study social communities and interactions among students and instructors [21–23] and to assess conceptual inventories developed for physics courses [24–29]. In general, these techniques are useful whenever connections between actors are of interest.

While prior work examining student understanding of the various notations used in quantum mechanics has been conducted, this has commonly been done only at individual institutions. Also, the ability to work and reason across multiple representations is important for success in QM. To better understand the ways in which students reason about expressions in multiple representations and to glean a more generalizable understanding of the same, we address the following research questions:

1. How can survey design and network analysis techniques be used in conjunction to efficiently collect and analyze data on students' conceptual connections between different QM expressions for many students at multiple institutions?
2. What can network analysis show about students' conceptual connections in spins-first courses?

We begin by laying out the work that has already been conducted in this space and providing a brief background on the relevant terms and concepts within network analysis that are used later. Then we discuss the design of our survey and the ways our networks were generated, before discussing the results of our analyses.

II. BACKGROUND

We begin our discussion of relevant prior research by reviewing work on student understanding of different mathematical representations used within quantum mechanics courses. We then provide a brief overview of prior work within PER using network analysis techniques, before laying out an overview of the specific network analysis concepts we are using by means of a toy model network.

A. Prior work with quantum mechanics representations

While much education research has been conducted at the boundary of physics and mathematics, there has been a focus more recently on the mathematics found in upper-division QM courses. This is especially true for the three mathematical notations commonly used to describe identical or analogous physical phenomena or concepts in upper-division QM: Dirac, wave function, and vector-matrix notations. Gire and Price examined all three notations from an expert perspective, noting the affordances and limitations of each for the purposes of computation [15]. For example, they found that students elected to use Dirac notation as a medium for coordinating expressions in other notations and attributed that preference to qualities of the notation, such as its compactness and symbolic support for computation. Schermerhorn *et al.* modified this framework to capture student preferences when calculating expectation values [14]; they found that both the compactness of a notation and its relative familiarity were the primary drivers for student preference. Wawro *et al.* similarly studied student judgments of vector-matrix and Dirac notations' suitability for particular applications and showed that a comprehensive understanding of both how similar expressions in these various notations interrelate and how to translate between them is crucial for a deep, cohesive understanding of QM [13]. Additionally, incorrect translation between wave function and Dirac expressions has been shown to lead to difficulties when developing models for calculating probabilities [30]. Because of the importance of this skill, instructional materials have been developed to assist students in working among and reasoning with multiple representations concurrently (e.g., simulation-based tutorials utilizing multiple types of graphs, integrals, and algebraic expressions [31]). We believe more work is necessary to understand how students reason about the ways in which expressions in these different notations interrelate, as this will allow better characterization of student thinking and ultimately allow for tailored pedagogy and instructional materials to improve this essential skill.

B. Prior education research studies using network analysis

Network analysis techniques have been developed and used to study topics as diverse as physical real-world infrastructure, neural networks, and social behaviors among groups. These techniques have recently also seen extensive use in both PER and studies in the broader education research literature more generally. Community detection and cluster analysis techniques have been used to study response groupings for various conceptual inventories [24–29] and, more recently, to interpret the results of Likert-style surveys [32]. Some work has been done to cluster students' knowledge, skills, and beliefs in the context of epistemological framing during computational physics [33]. These techniques have also been used extensively to study social communities and their various impacts, both among communities of educators [22,23] and of students [21,23,34], and to characterize how these social communities are affected by different active-learning pedagogies [35]. Recent work has even looked at how these communities have been affected by remote physics courses [36]. Though the techniques we use for community detection are much the same as those used in prior literature, the use of these techniques to study student interpretations of symbolic expressions is novel. An overview of these techniques is provided in Sec. III, and our specific utilization is detailed in Secs. IV and V.

III. NETWORK ANALYSIS PRIMER

Because network analysis is so broad and contains a multitude of terms that are still not commonplace in PER, a brief introduction to the specific terms and analysis methods that are applied to our data is merited. A toy model network, used to demonstrate different methods of our analysis, is shown in Fig. 1. Definitions of terms used are in Table I.

A. Community detection methods

There are many ways to detect community structure within networks, including by maximizing a network's modularity [37,38], agglomerative hierarchical clustering algorithms [39], and using edge betweennesses to continuously subdivide a network [40]. These various methods each have their strengths and weaknesses, which often manifest as trade-offs between computational speed, granularity of results, and level of confidence that can be ascribed

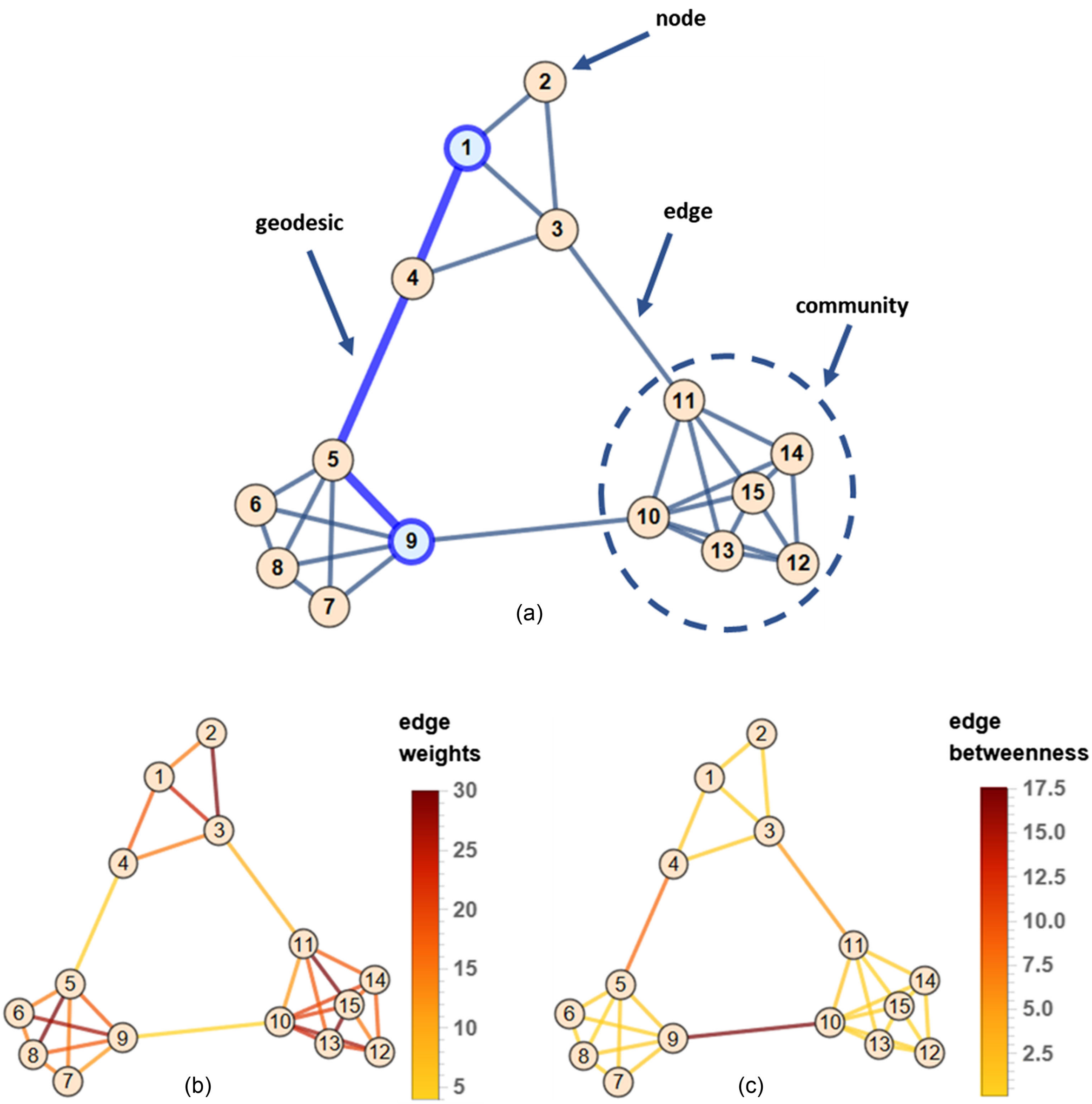


FIG. 1. A toy model network highlighting terms discussed in Table I. (a) Network highlighting the geodesic between nodes 1 and 9. (b) The same network, with the edges shaded by their edge weights; the edges lying within the three communities were made to have more weight than those that cross between communities. (c) The network with the edges shaded by their edge betweennesses; the edges lying between communities tend to have high betweennesses, while those within communities tend to have lower betweennesses.

TABLE I. Definitions and descriptions of relevant terms that will be used to discuss networks within this paper. For clarity, several of these terms are highlighted in Fig. 1 for the toy model network.

Term	Definition
Node	Individual objects that may or may not be connected to each other
Edge	Represents connection between nodes.
Edge weight	Corresponds to the strength of the connection between two nodes. If an edge has a large edge weight, that means that the connection between the two nodes on either end of that edge is particularly strong.
Community	Clusters of nodes that have been determined by some means to be more closely connected to each other than to the other nodes in the network.
Geodesic	The shortest possible path between two nodes in a network.
Edge betweenness	Determined by finding the geodesics between every pair of nodes in a network and counting the number that passes through the edge in question. For a weighted network like our toy network, this number is then divided by the edge's weight. Effectively, an edge has low betweenness if there are many alternative geodesics that can sidestep it, such as within a tightly knit community.

to the specific communities formed. Modularity maximization is relatively quick to compute but has limitations to its resolution of smaller communities [41]. Hierarchical clustering and edge betweenness-based methods both provide a much higher resolution of subcommunities by generating a hierarchical community structure, which can be visualized succinctly with a dendrogram (see Fig. 2 for the dendrogram for our toy model from Fig. 1). While hierarchical clustering algorithms typically start with every node disconnected and slowly cluster them together via some similarity (distance) measure, betweenness-based methods start with the larger network as a whole and separate it into smaller and smaller subdivisions. In terms of creating a dendrogram, one could view hierarchical clustering as building the dendrogram from the bottom up and betweenness-based methods as doing so from the top down. Both are computationally intensive for networks of even moderate size (for sparse graphs with n edges, completion times are, at best, $O(n^2)$ for hierarchical clustering and $O(n^3)$ for betweenness-based methods). While hierarchical clustering techniques are afforded some flexibility from their reliance on similarity measures between nodes—from which there are many options to choose—this also means that different metrics can provide different clusters without an obvious way to know if one is more correct than the others [42]. Additional drawbacks of hierarchical clustering are that even central members of communities can be left out of the communities they "should" belong within, and that it often leaves some peripheral members out as well [43]. Given the drawbacks of modularity maximization and hierarchical clustering, and the fact that the networks we will study are not overly large (at only $n = 20$ nodes), we decided to use edge betweenness to determine how our expressions were clustered.[1] It is notable that the hierarchical nature of these dendrograms leaves open the question of which number of communities is "best" for a given network. This can be visualized by slicing horizontally across a dendrogram at various vertical positions. For the dendrogram in Fig. 2, for example, we could say there are two communities [made up of (1–9) and (10–15)], or three communities [consisting of (1–4), (5–9) and (10–15)], or some other number, depending on where the cut is made. For example, the 5–9 community could be made from two subcommunities (5, 7, 8 and 6, 9), and/or the 10–15 community could be made from 14, 12, 15 and 11, 10, 13. This question and others, as well as how we go about answering them, are discussed in more detail in Sec. III B.

The way that our chosen algorithm works is by sequentially removing the edge within our network that has the highest betweenness. As can be seen from Fig. 1(c), edges that have the largest betweennesses are most likely to connect communities [44]. Once the edge with the highest betweenness is found, it is deleted, and the edge

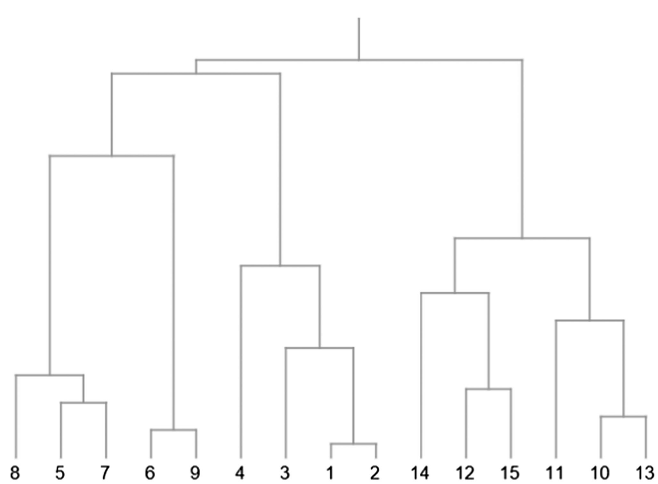


FIG. 2. Dendrogram showing the community structure of the toy model network from Fig. 1 found using the edge betweenness method. The three expected communities (1–4, 5–9, and 10–15) are clearly visible.

[1] For the curious reader, Ref. [42] provides an excellent review of various community detection methods and their respective affordances and limitations.

betweennesses of the resultant network are recalculated; this process is repeated until every edge in the network is removed and all vertices are fully disconnected. This process will tend to separate a network into progressively smaller communities, saving the most tight-knit for last; the result can be visualized with a dendrogram. In these dendrograms, the network begins with a single community, represented by a vertical line. As the algorithm runs its course, this single community splits into multiple branches containing fewer and fewer vertices. This is visualized by moving down the diagram until every edge is removed from the network and every vertex is separated from every other at the bottom of the dendrogram. While originally conceived as a method for community detection in unweighted networks, Newman extended this procedure to include networks with weighted edges, as is the case both for our toy model and for our survey data [45].

B. Determining community robustness

The relative height of a given vertical segment on these edge betweenness-based dendrograms is indicative of the number of edges that were removed between community divisions. Divisions with very little vertical space between them therefore occurred fairly close together during the community detection process. Before beginning analysis of these communities and drawing conclusions based on the order in which they are formed, some questions arise:

1. How robust is the community structure, in that small perturbations to the network would not produce a meaningfully different community structure?
2. How confident can we be about these communities?
3. Where should we "stop" along the vertical axis, to determine which community division or number of communities is "best"?

There are numerous possible methods to bring to bear to answer these questions, including determining which division has the largest modularity [44] and various bootstrapping procedures [42]. Due to the level of granularity and transparency afforded by the latter, we utilized a modified bootstrapping procedure to determine which community configuration represents the overall data most effectively and is robust and stable to perturbations.

In particular, we elected to use a technique based on statistical bootstrapping discussed by Fortunato [42] and Efron and Tibshirani [46] that was subsequently modified by Speirs *et al.* [47]. The basic idea is to resample from the pool of student responses to generate multiple hypothetical datasets. For a dataset of N student respondents, hypothetical datasets are created, each comprised of N responses drawn at random from the actual student responses. When creating a hypothetical dataset, some respondents' data will be selected multiple times, and others not at all—otherwise the process would simply reproduce the original dataset. A network is then created from this hypothetical dataset and the community detection algorithm is run. This process is then repeated many times to create many slightly different networks, with the resulting ensemble of dendrograms then being compared to see where significant deviations occur and what structure is common across all or most hypothetical datasets. In practice, determining this common structure means comparing the different communities formed across the bootstrapped networks to see which communities are consistently detected.

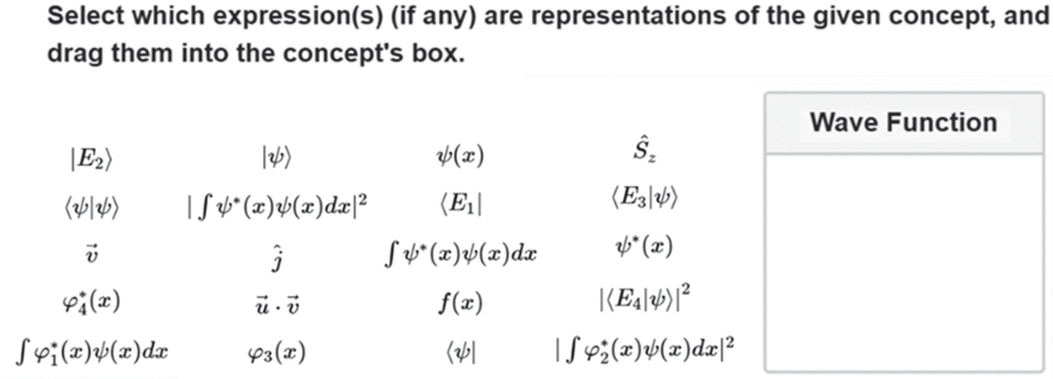
The modified bootstrapping procedure answers the questions listed above regarding how to determine "good" community divisions in the following ways: (i) It operates by introducing perturbations to the original network and drawing attention specifically to places where the perturbations did not drastically alter the community structure; (ii) the level of confidence afforded to any given community structure can be determined by the fraction of bootstrapped networks that agree on a given community structure. If 100% of bootstrapped networks have the same three-community division, it is reasonable to be very confident in that community division. Speirs *et al.* determined their confidence window to be for agreement among >60% of bootstrapped networks [47]; (iii) whichever community division(s) has the greatest agreement among the bootstrapped networks can be reasonably treated as the "best" community division(s). This process is described in more detail in Sec. V A as we apply it to our data.

IV. METHODS

The discussion of our research methods begins with a discussion regarding our survey design and implementation, before then describing the creation of our networks from survey data.

A. Survey design

In an effort to be able to make more generalizable claims about student understanding of the representations used in upper-division QM, we developed and administered an online survey with two primary goals in mind: to collect and analyze responses from many students across multiple institutions in a way that would scale efficiently and to create a dataset that allows for analysis of students' conceptual understanding of mathematical expressions commonly used in QM—particularly those used to express probability concepts. Because both wave function and Dirac notations are used extensively in upper-division QM courses, and because there are equivalent expressions that look quite different between the two notations (e.g., $\langle E_n|\psi\rangle$ and $\int \varphi_n^*(x)\psi(x)dx$) as well as similar-looking expressions that represent concepts with very subtle distinctions (e.g., $|E_n\rangle$ & $\langle E_n|$), we decided to focus on the conceptual interpretations that various commonly used expressions shared.



(a)

Concepts		
Vector	Wave function	Eigenvector
Quantum state	Unit Vector	Probability amplitude
Inner product	Basis Vector	Probability
Dot product	Eigenstate	

(b)

FIG. 3. (a) Example of a prompt in the online survey administered to students. (b) Table showing the different concepts for which they were tasked with selecting expressions.

The first goal was addressed by reducing the number of free-response text entry questions as much as possible, both to reduce participant attrition and to help our analysis scale well to a large participant pool. This meant that the second goal—gleaning students' conceptual connections between expressions—would need to be accomplished without explicitly analyzing student reasoning. To this end, the survey tasks were designed as sorting tasks, where the students were presented with a list of expressions commonly used in upper-division QM courses [Fig. 3(a)] and a single quantum mechanical concept. Students were tasked with selecting all of the expressions in the list that they felt represented that concept. In all, the survey consisted of 11 different concepts [Fig. 3(b)]. Each participant was presented with one concept at a time [as shown in Fig. 3(a)] to avoid having too much information on the screen at one time. The expressions and concepts were selected to balance the sufficient breadth of each while avoiding overloading participants with too many options and survey fatigue. The final list of expressions was settled on after a pilot test of the survey. The concepts were chosen to cover most of the normative interpretations for the different expressions selected. Additionally, this survey was used as a part of a larger research effort to investigate student understanding of expressions for probability concepts in QM; this also affected the selection of expressions and concepts.

This survey was given to students in upper-division QM courses at six different institutions ($N = 139$); all courses were taught using a spins-first textbook ([17] or [18]). As discussed in Sec. II A, these courses typically begin by using the results of the Stern-Gerlach experiment to motivate the treatment of quantum states as state vectors, often represented by Dirac notation bras and kets. After some time studying systems with discrete measurement outcomes in Dirac notation, including the time evolution of these systems, these courses eventually transition to studying continuous systems, connecting the Dirac state vectors to their associated wave functions. The survey was distributed near the end of the course after students had worked extensively with both Dirac and wave function notations. Courses using the spins-first approach were chosen for this study due to the increasing prevalence of this curricular style as compared to more traditional wave functions-first courses, which typically use the text by Griffiths [16]. It was also a population of convenience, as the institution at which this survey was developed and piloted used a spins-first approach. We also suspect that due to the increased and early focus on Dirac notation in these courses, it was possible that expressions in Dirac notation would be more commonly selected. As such, we viewed this population as a useful one for testing this methodology before extending it to study participants from wave functions-first curricula.

As a result of the survey design, student responses are entirely relational, forming expression-concept pairs and expression-expression pairs. An expression-concept pair is formed between an expression and a concept when the expression is dragged into the concept's box, and an expression-expression pair is formed when two expressions are selected simultaneously for the same concept, i.e., dragged into the same box. Expression-concept pairs, though important, are not the focus of this study. We focus on expression-expression pairs as representative of a shared conceptual interpretation on the part of the students in Sec. V B. We also focus on the concepts for which students selected these expression-expression pairs as that helps to explain what those shared conceptual interpretations are. Given the nature of the survey responses and our interest in how and whether students view these expressions as conceptually connected, we decided to use network analysis techniques to analyze our survey responses. In particular, we implemented the edge betweenness community detection method discussed in Sec. III.

B. Creating our network

To turn our survey results into a network for analysis of how these expressions are related conceptually, we first collapsed individual students' responses into their own networks. In these networks, the nodes were the 20 different expressions provided on the survey; a connection was placed between two nodes if the student declared those two expressions as representative of the same concept, i.e., placed those two expressions in the same concept box. Here we made a choice that even if a student selected the same pair of expressions for multiple concepts, each pair was only counted once per student, resulting in "unweighted" individual student response networks. This unweighting was done to avoid overweighting any connections that may be due to similar concepts given on the survey. For example, if a student selected $\langle E_1|$ and $|E_2\rangle$ simultaneously for the concepts "vector," "basis vector," and "eigenvector," those two expressions would be connected by an edge of weight 1 for that student (rather than 3). While we wanted the granularity of these different types of vectors for other analyses in our broader study, we did not want to overweight the strength of those connections for this investigation simply because there were more distinct vector concepts on the survey than there were variations of other concepts (e.g., those related to probability or quantum states) while constructing these students' individual networks. These links then served to connect expressions that students believed could represent the same concept.

For N respondents, this process resulted in N unweighted networks of 20 nodes each. These N networks were then superimposed to generate the full weighted network (Fig. 4), with a maximum possible edge weight of N if all respondents selected the two expressions connected by that edge simultaneously at least once on the survey. This network is rather complicated, with myriad low-weight edges that make it difficult to interpret. One way to cut through this visual clutter and help meaningful information rise to the surface is to find the clusters or communities within the larger network as a whole. To this end, we implemented the edge betweenness community detection method discussed in Sec. III.

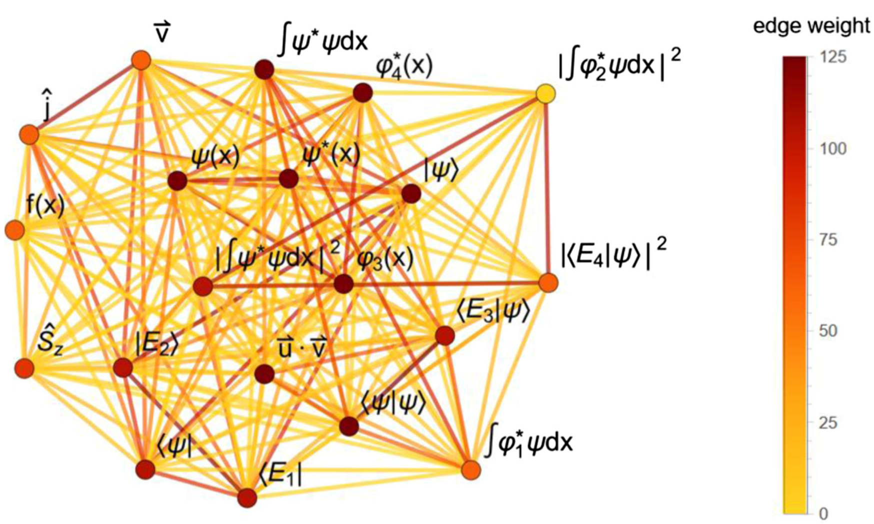


FIG. 4. The expression-concept network generated from 139 student survey respondents. The nodes represent the different expressions on the survey. Edges between nodes show that students selected those two expressions simultaneously for at least one concept. The edge weights represent the number of students who used the two expressions simultaneously and is shown by the shading applied to the edges. The shading of the nodes is not a part of our analysis here but is representative of the nodes' degree (the number and weight of the edges connected to each node).

V. RESULTS AND DISCUSSION

We begin the discussion of our results by first presenting the details of the community detection process. We then discuss the reliability of the detected communities as well as the communities' implications for student understanding of mathematical representations within this context.

A. Running the betweenness algorithm and determining its limits

The dendrogram generated by the edge betweenness method for our network is shown in Fig. 5. Cutting horizontally across this dendrogram at any point is representative of a snapshot of the betweenness algorithm—it represents the set of communities that exist at any given point during the procedure.

As discussed in Sec. III B, community divisions that are separated by relatively small vertical segments on the dendrogram occurred relatively close together during the betweenness algorithm. Given the importance of the order in which these divisions occur for our analyses, a determination of the community divisions that can be considered robust is of great importance. To this end, we implemented the statistical bootstrapping-based method discussed in Sec. III B to determine the robustness of these community divisions. In particular, we ran the edge betweenness community detection algorithm on 1000 bootstrapped networks. This number was chosen to balance computation time and accuracy: the bootstrapping procedure was run several times, and 1000 networks were found to generally produce consistent results while taking an acceptable amount of time to compute. Figure 6 shows the number of dendrograms that have identical community structures at a given number of communities (level of the dendrogram). Each bar on this plot represents a specific number of communities—from 1 community (with every node included) to 20 (with every node separated into its own communities at the end of the betweenness procedure), as indexed on the horizontal axis. Each differently shaded segment of a bar represents a subset of the hypothetical (bootstrapped) networks with both (a) the same number of communities and (b) the exact same set(s) of communities. Note that the bar on either end of the plot has only one segment: there should only be one possible structure with 1 community—with all nodes connected (the top of the dendrogram)—and one with 20 communities—with each node as its own community (the bottom of the dendrogram). One interesting aspect of this figure can be seen by observing the second and third bars on this plot. The second bar shows that the 1000 bootstrapped networks showcase three different initial divisions of the networks into two communities: 525 networks result in one common pair of communities, 343 in another distinct pair, and 132 divide a third way. Upon dividing once more and reaching a three-community division of the bootstrapped networks, however, more than 750 of the bootstrapped networks share identical communities. This consolidation effect is due largely to variability in the order of the early community divisions, as the first two divisions visible in Fig. 5 frequently swap their order among the bootstrapped networks. The numbers of communities for which there are very few different community structures, evidenced by a small number of stacked segments, and/or one dominant community structure, evidenced by one much larger segment, are thus indicative of high agreement among bootstrapped networks at that level of their respective dendrograms, and vice versa. With this in mind, Fig. 6 shows where there is high and low agreement among the

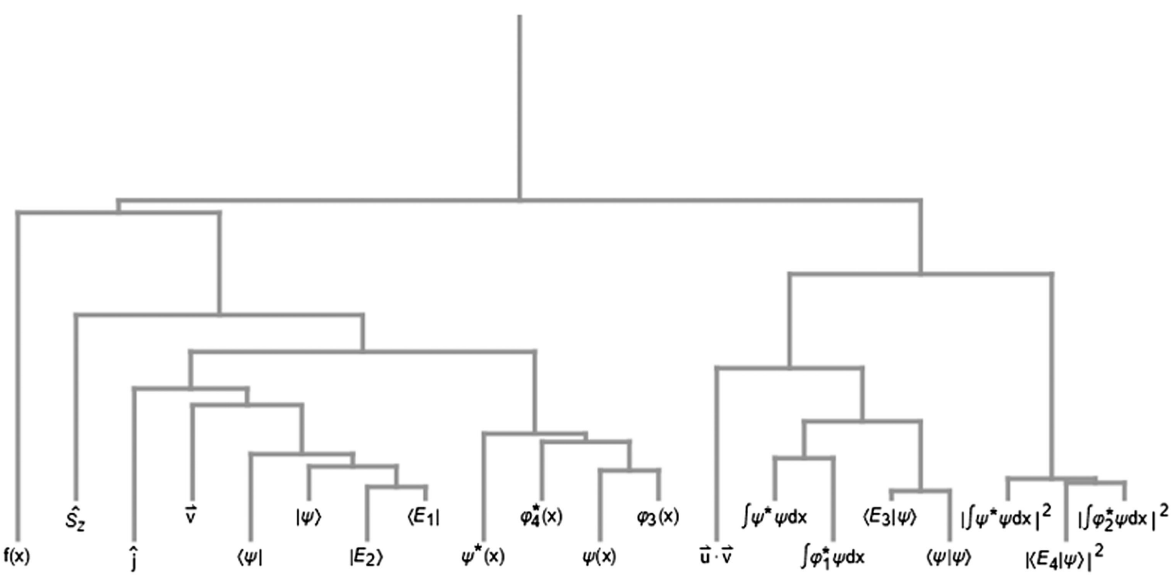


FIG. 5. Dendrogram showing the community structure of the network in Fig. 4 after implementing the edge betweenness community detection algorithm.

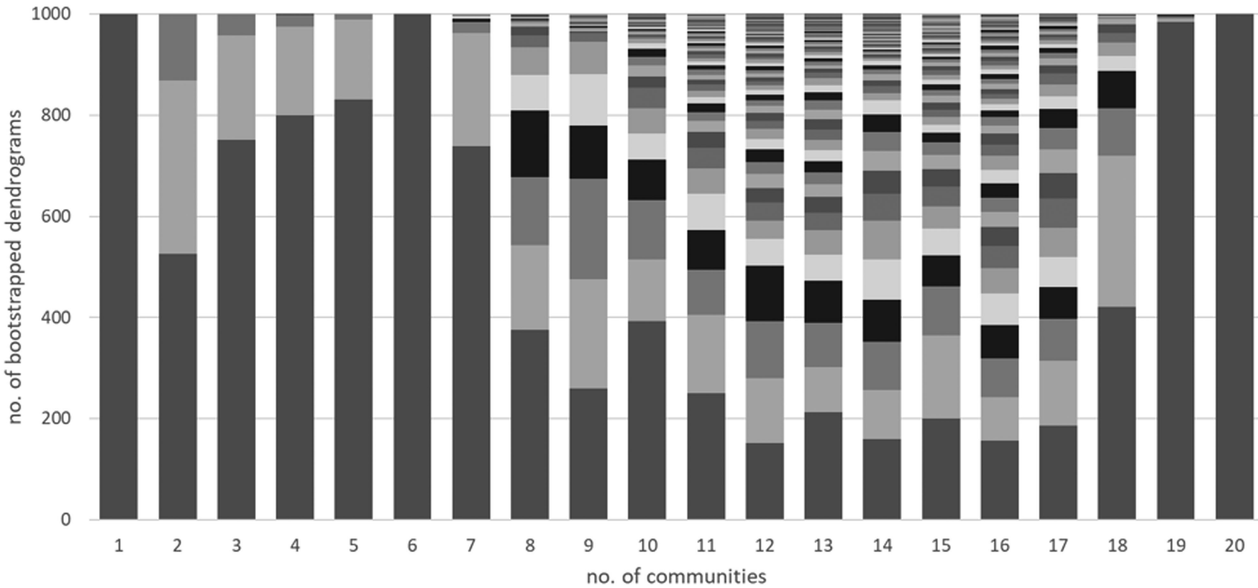


FIG. 6. Stacked bar chart showing the relative proportions of different community structures at each level across the bootstrapped dendrograms. Each individual bar represents a specific community structure that is shared among some portion of the bootstrapped networks at a given level of the edge betweenness algorithm.

bootstrapped dendrograms, and thus we can look for the level of the community detection algorithm for which the community structure is most stable under perturbations. As discussed in Sec. III B, earlier work found agreement among >60% of bootstrapped networks to be enough to be confident as to the existence of a robust community [47].

As can be seen in Fig. 6, some variation occurs within the 2–5 community range (often with a single community structure being represented in >60% of the bootstrapped networks), but 100% of the bootstrapped networks have an identical community structure once they are broken into six communities. Thus, we primarily focus on these six communities. For more than seven communities, the stability swiftly devolves, before ultimately agreeing strongly again once nearly every vertex is separated for all of the bootstrapped networks (as should be expected: there is only one way for 20 nodes to be separated into 20 communities). One finding to take from this is that while there is minor variability in the relative order of the first four divisions of our network (the four highest splits on the dendrogram in Fig. 5), we have high confidence that the division of the network into the six communities seen in Fig. 7 happens prior to any of the divisions below it on the dendrogram. This suggests that these initial six communities are highly stable.

B. Interpreting community structure

Separating our initial network into the first six communities determined by the edge betweenness algorithm gives us the network seen in Fig. 7. Two of the communities consist of $\hat{S}_z$ and $f(x)$ individually (labeled SZ and FX in Fig. 7, respectively), which shows that students did not

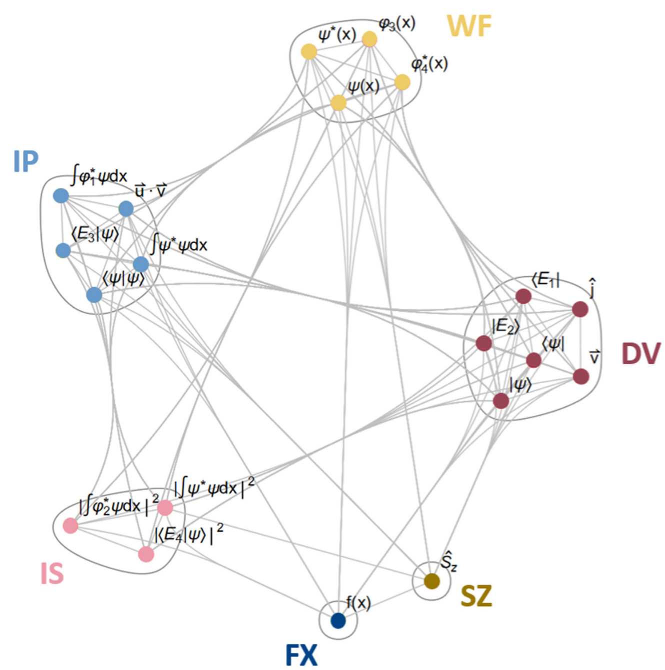


FIG. 7. The network built from student survey responses, as in Fig. 4, grouped into the six stable communities as determined by the bootstrapping procedure shown in Fig. 6. FX contains only the generic function $f(x)$; SZ contains only the spin-Z operator $\hat{S}_z$; DV contains Dirac bras, kets, and generic vector expressions $\vec{v}$ and $\hat{j}$; WF contains the wave function and its conjugate [$\psi(x)$ and $\psi^*(x)$], as well as the wave function and conjugate wave function expressions for eigenstates ($\varphi_3(x)$ and $\varphi_4^*(x)$); IP contains inner product expressions, including a generic dot product in both Dirac and wave function notation; IS contains complex squares of inner products in both notations.

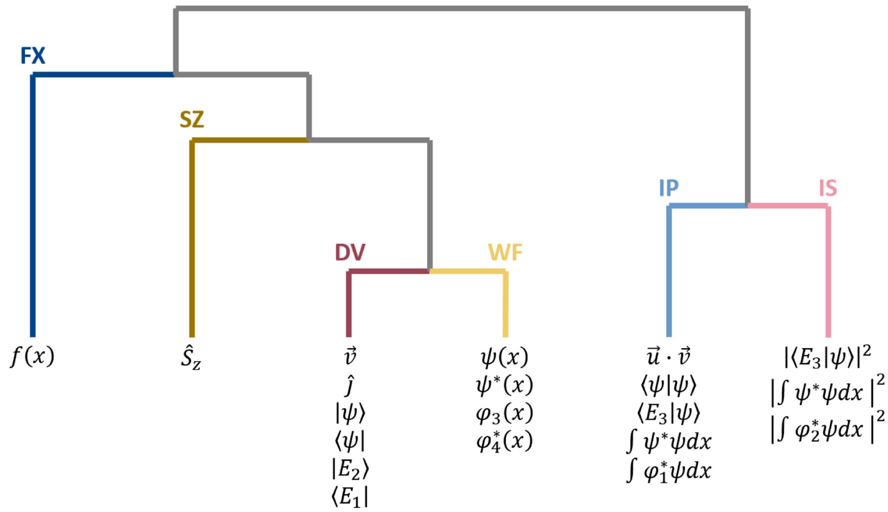


FIG. 8. Simplified dendrogram showing the stable communities from the bootstrapping procedure, color coded and labeled to match Fig. 7.

think of either of these expressions as especially conceptually similar to the other 18 expressions. The Dirac bras, kets, and generic vector expressions $\vec{v}$ and $\hat{j}$ were found to form their own community (DV), as were the wave function expressions (WF). The expressions for inner products, including a generic dot product $\vec{u}\cdot\vec{v}$, were also separated into a community (IP), as were the expressions for the complex squares of inner products (IS).

IP and IS's separation suggests a conceptual distinction between inner products and the squares of inner products (in QM, this is often a distinction between probability amplitudes and probabilities). Similarly, the separation of communities WF and DV suggests a meaningful distinction between wave functions and Dirac bras and kets and generic vector expressions.

A closer look at the earlier divisions of the network seen in the dendrogram in Fig. 5 reveals two larger-grain-size communities: that of IP + IS and that of WF + DV. A simplified version of this dendrogram is shown in Fig. 8, where the divisions beyond the six-community 100% confidence limit determined by the bootstrapping procedure are eliminated. The combination of this larger-scale (IP + IS and WF + DV) community structure and our survey design leads us to the conclusion that the expressions in WF and DV are viewed as conceptually more similar to each other than to the remaining expressions; the same is true for the expressions in IP and IS. While the edges connecting these subcommunities together represent conceptual connections for our students, the connections *within* each subcommunity are stronger than those connecting the subcommunities to each other. In other words, the expressions within the WF subcommunity are more tightly knit to themselves than to the expressions within the DV subcommunity, but those two subcommunities are still connected more closely (into a larger WF + DV community) than they are to the other expressions in the survey. A similar statement can be made about the IP, IS, and IP + IS communities.

For the larger WF + DV community, we can also look at the conceptual makeup of these connections to obtain a deeper understanding of which specific concepts connect these expressions. To this end, we separate the expressions within this larger group into three different groups, by notational style: one for Dirac bras and kets ($|\psi\rangle$, $\langle\psi|$, $|E_2\rangle$, and $\langle E_1|$), one for the generic individual vector expressions ($\vec{v}$ and $\hat{j}$), and one for the wave function expressions [$\psi(x)$, $\psi^*(x)$, $\varphi_3(x)$, and $\varphi_4^*(x)$]. We then look at the concepts that connect these different types of expressions together according to the students. These expressions are ultimately connected because participants selected them as representing the same concepts, and thus we can examine pairs of expressions along this axis by looking at *which* concepts students connected them with. To this end, as we did with the expressions themselves, we separate the concepts used by students to connect these types of expressions into three camps: vectors ("vector," "eigenvector," "unit vector," and "basis vector"), quantum states ("quantum state" and "eigenstate"), and wave functions (just "wave function"). We then break down the proportions of each type of concept that connected the various types of expressions, both within a given expression type or between different expression types. We focus on a subset of these conceptual connections between expression types, shown in Fig. 9. These specific pairs were selected for their illustrative nature; WF-vector and vector-vector were excluded because these connections were either very uncommon or uninteresting, respectively.

By examining these conceptual breakdowns, we can see that the concepts for which students used both a Dirac bra or ket ($|\psi\rangle$, $\langle\psi|$, $|E_2\rangle$, or $\langle E_1|$) and a generic vector

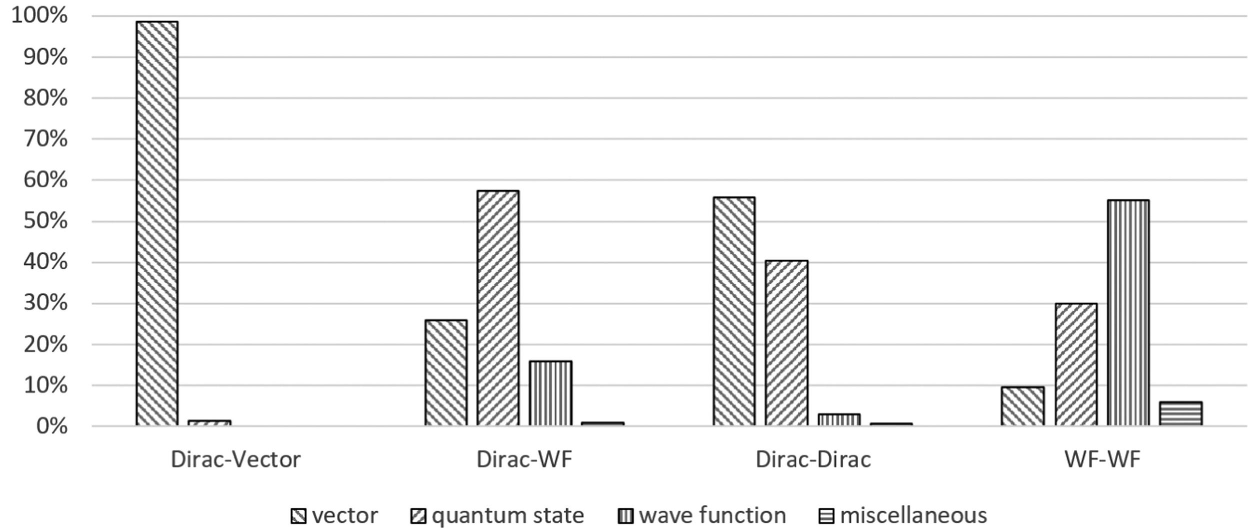


FIG. 9. Histogram comparing the types of concepts used to link Dirac bras and kets with generic vector expressions and wave function expressions, as well as the types of concepts that connect Dirac bra and ket expressions to each other and wave function expressions to each other.

expression ($\vec{v}$ or $\hat{j}$) (i.e., concepts represented by the connections between those two subsets of nodes) are almost entirely vector concepts. This is noticeably distinct from the distribution of concepts that consistently linked the Dirac bras and kets to the wave function expressions on the survey ($\psi(x)$, $\psi^*(x)$, $\varphi_3(x)$, and $\varphi_4^*(x)$), more than half of which were those related to quantum states. Taken in combination with the stable and distinct WF and DV communities as seen in Figs. 7 and 8 (i.e., that these vector-Dirac connections were stronger than the Dirac-wave function connections), this suggests that the vectorlike identity of the Dirac bras and kets was stronger for these students than either their quantum state- or wave function-like conceptualizations. Also, the edges connecting the Dirac bras and kets to each other are split between vector and quantum state concepts, while the connections between wave function expressions are split between wave function and quantum state concepts. This would seem to suggest that while Dirac bras and kets share a strong identity as representing vectorlike concepts and wave

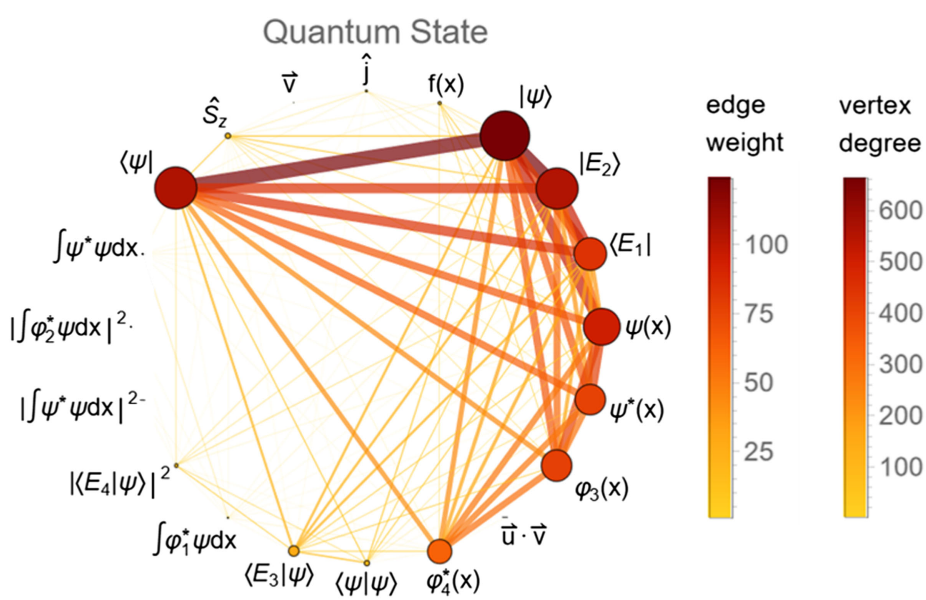


FIG. 10. Network showing spins-first students' connections between different expressions when prompted to select expressions representative of "quantum state." Edges are sized proportional to and colored according to their weight (i.e., the number of students who selected the pair of expressions for this concept). Nodes are similarly sized proportional to and colored according to their vertex degree (i.e., the sum of the weights of all edges that connect to them).

function expressions share a strong identity as representing wave function concepts, both appear to be recognized as representing quantum state concepts. This can be seen more explicitly by looking at a network composed of the expressions students used simultaneously to represent the *quantum state* concept, shown in Fig. 10. This network displays strong connectivity between bras, kets, functions, and conjugate functions, providing evidence that all of these expressions are treated by these students as representing quantum states, despite the more siloed interpretation of bras and kets representing vectors (and not so much wave functions) and of wave functions representing wave functions (and not so much vectors).

These findings are to be expected from the curricular focus of the courses in which these students were enrolled. Spins-first quantum mechanics courses begin by introducing Dirac notation in two-dimensional spin-$1/2$ bases and lean heavily on familiar vector interpretations to help students understand the mathematical operations at play. Probability amplitude inner products at the beginning of the course are very much treated as geometric dot product projections, with state vectors having components along the different eigenstates' "directions." This analysis suggests that these curricular goals have been successful in getting these students to think about Dirac bras and kets as simultaneously representative of both vectors and quantum states. That $\vec{u}\cdot\vec{v}$ is included within the IP community likewise suggests that these students see this connection between inner products (both as Dirac brackets and wave function integrals) as sharing conceptual backing with dot products—another common instructional goal within these courses.

VI. CONCLUSIONS AND FUTURE WORK

Prior studies have shown that each notation used in QM has certain aspects that make it more suited for certain tasks [14,15] and that the ability to effectively and efficiently translate between notations is a crucial skill for students to develop [13]. In fact, incorrectly translating between notations has been shown to hinder students' ability to develop models for calculating probabilities [30]. In our study, network analysis techniques were used to investigate the strengths of students' conceptual connections between common expressions in upper-division quantum mechanics. We found that Dirac bras and kets and their associated wave functions both have distinct shared conceptual identities as vectors and wave functions, respectively, but they also both represent quantum state concepts to the participants in our study. It is possible that this shared quantum state identity aids students by serving as a touchstone when translating from one notation to the other or that this could serve as a starting point for curriculum or pedagogy aimed at improving this skill.

We also found from our community detection analysis that Dirac bras and kets were considered more conceptually similar to generic vector expressions than they were to their equivalent wave function representations. This is clear from their separation into the DV and WF communities. An implication of this result is that Dirac bras and kets bear a strong association with vectorlike concepts—even more than with conceptualizations they share with wave functions, such as both being representations for quantum states. The strength of this association with vectorlike concepts is perhaps to be expected due to the curricular structure of these courses, as all of these spins-first courses begin by first drawing attention to the discrete vectorlike nature of bras and kets before eventually connecting them to continuous wave function interpretations. While a lack of resolution prevents us from looking closer at the connections between the expressions within the inner product community (IP in Fig. 7), the result that the generic dot product remains a part of that community suggests a seemingly dot productlike understanding of even the wave function inner product integrals. This is an encouraging finding, as the conceptual connections between discrete and continuous inner products are important to develop and can often prove elusive.

More broadly, we have shown that with the novel use of online survey design and network analysis techniques, an investigation of student understanding of mathematical expressions in quantum mechanics and their interrelated conceptual interpretations is feasible for a large number of students at multiple institutions. We believe that the scalability of this methodology can allow for greater generalizability of findings. We also believe that these methods could be used to study conceptual interpretations of other expressions, both in QM and beyond. This study focused on expressions relating to inner products, but there are a multitude of expressions within physics that share multiple conceptual interpretations—these methods could prove useful in efficiently studying the ways that large numbers of students relate such expressions to each other and how strongly they relate them.

Furthermore, while much work has been done with network analysis and community detection algorithms within the PER community, there has been relatively little work done in examining the relative stability and robustness of the communities that have been studied. As illustrated in Fig. 6, taking small grain-size community structures at face value can be fraught with potential errors. As the resolution of any community detection algorithm is limited, the use of bootstrapping or similar techniques to help determine the resolution of a given community structure seems to be a productive approach for making community-based claims about network structure.

Our next step is to extend this work to include students enrolled in the more traditional wave functions-first courses. While many of the broader findings from community detection for respondents in wave functions-first courses may be similar to those seen here, we suspect that there may be quite different findings when it comes to the vectorlike interpretations of the more quantum mechanical expressions due to the relative lack of focus on linear algebraic interpretations in such courses. Similarly, a comparison of the expressions chosen to represent individual concepts may be of interest, particularly vectorlike concepts due to the lessened focus on linear algebra-based reasoning at the beginning of wave functions-first courses.

Another possible extension of this work would be an in-depth qualitative study of the specific conceptual interpretations students deemed these various expressions to represent. A natural framework to use for this type of analysis would be that of Sherin's symbolic forms [48] or its recent extension, symbolic blending [49]. Applying either or both of these frameworks could provide a deeper understanding of these expressions' conceptual interpretations.

ACKNOWLEDGMENTS

This work is supported in part by the National Science Foundation under Grant No. PHY-1912087. We would also like to thank our reviewers who all gave extremely constructive feedback and made this paper undeniably stronger.

[1] O. Uhden, R. Karam, M. Pietrocola, and G. Pospiech, Modelling mathematical reasoning in physics education, Sci. Educ. (Dordrecht) **21**, 485 (2012).
[2] C. Singh, Student understanding of quantum mechanics, Am. J. Phys. **69**, 885 (2001).
[3] C. Singh and E. Marshman, Review of student difficulties in upper-level quantum mechanics, Phys. Rev. ST Phys. Educ. Res. **11**, 020117 (2015).
[4] C. Singh, Student understanding of quantum mechanics at the beginning of graduate instruction, Am. J. Phys. **76**, 277 (2008).
[5] G. Passante, B. P. Schermerhorn, S. J. Pollock, and H. R. Sadaghiani, Time evolution in quantum systems: A closer look at student understanding, Eur. J. Phys. **41**, 015705 (2020).
[6] P. J. Emigh, G. Passante, and P. S. Shaffer, Student understanding of time dependence in quantum mechanics, Phys. Rev. ST Phys. Educ. Res. **11**, 020112 (2015).
[7] G. Karakok, Making connections among representations of eigenvector: What sort of a beast is it?, ZDM Int. J. Math. Educ. **51**, 1141 (2019).
[8] M. Wawro, K. Watson, and M. Zandieh, Student understanding of linear combinations of eigenvectors, ZDM Int. J. Math. Educ. **51**, 1111 (2019).
[9] M. O. J. Thomas and S. Stewart, Eigenvalues and eigenvectors: Embodied, symbolic and formal thinking, Math. Educ. Res. J. **23**, 275 (2011).
[10] M. Wawro, K. Watson, and W. Christensen, Meta-representational competence with linear algebra in quantum mechanics, in *Proceedings of CERME 10* (2017).
[11] A. Pina, Z. Topdemir, and J. R. Thompson, Student understanding of eigenvalue equations in quantum mechanics: Symbolic forms, in *Proceedings of 25th RUME* (2023).
[12] K. S. Serbin and M. Wawro, The inextricability of students' mathematical and physical reasoning in quantum mechanics problems, Int. J. Res. Undergrad. Math. Educ. **10**, 57 (2022).
[13] M. Wawro, K. Watson, and W. Christensen, Students' metarepresentational competence with matrix notation and Dirac notation in quantum mechanics, Phys. Rev. Phys. Educ. Res. **16**, 020112 (2020).
[14] B. P. Schermerhorn, G. Passante, H. R. Sadaghiani, and S. J. Pollock, Exploring student preferences when calculating expectation values using a computational features framework, Phys. Rev. Phys. Educ. Res. **15**, 020144 (2019).
[15] E. Gire and E. Price, Structural features of algebraic quantum notations, Phys. Rev. ST Phys. Educ. Res. **11**, 020109 (2015).
[16] D. J. Griffiths, *Introduction to Quantum Mechanics*, 1st ed. (Pearson, Upper Saddle River, NJ, 1995).
[17] D. H. McIntyre, *Quantum Mechanics: A Paradigms Approach* (Pearson, San Francisco, CA, 2012).
[18] J. S. Townsend, *A Modern Approach to Quantum Mechanics* (University Science Books, Mill Valley, CA, 2000).
[19] J. Stewart, J. Hansen, and E. Burkholder, Visualizing and predicting the path to an undergraduate physics degree at two different institutions, Phys. Rev. Phys. Educ. Res. **18**, 020117 (2022).
[20] M. E. J. Newman, *Networks: An Introduction*, 1st ed. (Oxford University Press, New York, 2010).
[21] S. Thomas, Ties that bind: A social network approach to understanding student integration and persistence, J. Higher Educ. **71**, 591 (2000).
[22] P. S. Smith, M. L. Hayes, and K. M. Lyons, The ecology of instructional teacher leadership, J. Math. Behav. **46**, 267 (2017).
[23] M. Hopkins, D. Ozimek, and T. M. Sweet, Mathematics coaching and instructional reform: Individual and collective change, J. Math. Behav. **46**, 215 (2017).
[24] J. Wells, R. Henderson, J. Stewart, G. Stewart, J. Yang, and A. Traxler, Exploring the structure of misconceptions in the Force Concept Inventory with modified module analysis, Phys. Rev. Phys. Educ. Res. **15**, 020122 (2019).
[25] J. Wells, R. Henderson, A. Traxler, P. Miller, and J. Stewart, Exploring the structure of misconceptions in the Force and Motion Conceptual Evaluation with modified module analysis, Phys. Rev. Phys. Educ. Res. **16**, 010121 (2020).
[26] 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).

[27] C. Wheatley, J. Wells, R. Henderson, and J. Stewart, Applying module analysis to the conceptual survey of electricity and magnetism, Phys. Rev. Phys. Educ. Res. **17**, 010102 (2021).

[28] 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).

[29] 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).

[30] T. Wan, P. J. Emigh, and P. S. Shaffer, Investigating how students relate inner products and quantum probabilities, Phys. Rev. Phys. Educ. Res. **15**, 010117 (2019).

[31] A. Kohnle and G. Passante, Characterizing representational learning: A combined simulation and tutorial on perturbation theory, Phys. Rev. Phys. Educ. Res. **13**, 020131 (2017).

[32] R. P. Dalka, D. Sachmpazidi, C. Henderson, and J. P. Zwolak, Network analysis approach to Likert-style surveys, Phys. Rev. Phys. Educ. Res. **18**, 020113 (2022).

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

[34] 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).

[35] 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).

[36] M. Sundstrom, A. Schang, A. B. Heim, and N. G. Holmes, Understanding interaction network formation across instructional contexts in remote physics courses, Phys. Rev. Phys. Educ. Res. **18**, 020141 (2022).

[37] M. E. J. Newman, Finding community structure in networks using the eigenvectors of matrices, Phys. Rev. E **74**, 036104 (2006).

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

[39] R. P. Springuel, M. C. Wittmann, and J. R. Thompson, Applying clustering to statistical analysis of student reasoning about two-dimensional kinematics, Phys. Rev. ST Phys. Educ. Res. **3**, 020107 (2007).

[40] M. E. J. Newman and M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E **69**, 026113 (2004).

[41] S. Fortunato and M. Barthélemy, Resolution limit in community detection, Proc. Natl. Acad. Sci. U.S.A. **104**, 36 (2007).

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

[43] M. E. J. Newman, Detecting community structure in networks, Eur. Phys. J. B **38**, 321 (2004).

[44] M. E. J. Newman and M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E **69**, 026113 (2004).

[45] M. E. J. Newman, Analysis of weighted networks, Phys. Rev. E **70**, 056131 (2004).

[46] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993).

[47] J. C. Speirs, M. R. Stetzer, and B. A. Lindsey, Utilizing network analysis to explore student qualitative inferential reasoning chains, Phys. Rev. Phys. Educ. Res. **20**, 010147 (2024).

[48] B. L. Sherin, How students understand physics equations, Cognit. Instr. **19**, 479 (2001).

[49] B. P. Schermerhorn and J. R. Thompson, Making context explicit in equation construction and interpretation: Symbolic blending, Phys. Rev. Phys. Educ. Res. **19**, 020149 (2023).