

Characterizing faculty online learning community interactions using social network analysis

Emily Bolger¹, Marius Nwobi^{1,2}, and Marcos D. Caballero^{1,3,4}

¹*Department of Computational Mathematics, Science, and Engineering, Michigan State University, East Lansing, Michigan 48824, USA*

²*Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, USA*

³*Department of Physics and Astronomy and CREATE for STEM Institute, Michigan State University, East Lansing, Michigan 48824, USA*

⁴*Department of Physics and Center for Computing in Science Education, University of Oslo, 0315 Oslo, Norway*



(Received 2 July 2024; revised 3 August 2025; accepted 16 September 2025; published 21 October 2025)

The Partnership for Integration of Computation into Undergraduate Physics (PICUP) was founded in the mid-2010s to assist educators with the challenges of integrating computation into physics curricula. In addition to in-person trainings and hosted educational materials, PICUP uses a Slack Workspace to continue collaboration and discussion off-line. In this work, we use social network analysis (SNA) to study the communication patterns of PICUP and assess if PICUP is meeting their goals in the Slack environment. Through our analysis, we discuss PICUP's community structure and define a conceptual framework to evaluate if the goals are being met through SNA metrics. We present a comprehensive analysis of eight channels in the Slack Workspace using various SNA metrics, identifying three distinct levels of user engagement. We conclude with implications for PICUP and provide recommendations for the community.

DOI: [10.1103/n3yf-5h29](https://doi.org/10.1103/n3yf-5h29)

I. INTRODUCTION

In the past two decades, there has been a push for more students to experience and learn computing at various stages of their career—CS 10 K Project, CSforALL. These arguments range from the desire for a more computationally literate population to preparing more individuals for modern careers in science, technology, engineering, and mathematics. Related, there has been a push for integrating computation into specific disciplinary spaces to strengthen students' computing skills outside of a traditional computer science classroom [1–3]. This includes reports in the physics community such as the *AAPT Recommendations for Computational Physics in the Undergraduate Physics Curriculum* [4] and *Phys21: Preparing Physics Students for 21st-Century Careers* [5]. Compared to other disciplines, it appears that physics is leading the way [6,7]. Regardless, instructors face numerous barriers when integrating computation into established courses, including expectations of content coverage, lack of instructor time, student resistance, and department norms [3,8].

The Partnership for Integration of Computation into Undergraduate Physics (PICUP) [9] was started in the mid-2010s to assist educators with the challenges that arise from designing computing curricula for a noncomputing classroom. The community supports its members in a variety of ways. From hosting in-person workshops to online webinars to example exercise sets on their webpage to a Slack Workspace, instructors have a variety of opportunities to engage with material and construct a community of colleagues with similar goals. Since its inception, the community has served hundreds of faculty through its workshops and materials and has produced more than 15 publications.

Learning communities, particularly in the context of higher education, are nothing new. Baker [10] describes learning communities as a group “who have a clear sense of membership, common goals, and opportunity for extensive face-to-face interaction [11].” Often, these learning communities are situated in the communities of practice (CoP) framework, which is defined by three components: joint enterprise, mutual engagement, and shared repertoire [11–14]. While the definition of CoP is broad, there is an implied structure of different levels of participants, with the goal of periphery participants transitioning to core, active participants through engaging with the community.

For faculty, learning communities can be a successful mechanism for instructor professional development (PD) when built through a sustained community of support and learning [14–16]. With the rapid development of

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

technology and the ability to communicate remotely, the use of online communities as continued engagement or the sole form of PD is becoming more common as it allows educators to participate who might not be able to otherwise because of time and travel constraints [13,17–19]. Defined by Dancy *et al.* [13], faculty online learning communities (FOLCs) build off existing tenets of strictly in-person faculty learning communities (FLCs). The authors argue the following characteristics are some of the key components of FOLCs: periodic, potentially remote meetings over a set time; a supportive community of peers and experts united by a common interest; and the fostering of lifelong reflective instructors [11].

A variety of methodologies have been used to assess which strategies are most successful for sustained change in FLCs, FOLCs, and other online communities [11]. Commonly, researchers conduct a pre/post survey analysis and interviews with participants about their experience, performing qualitative coding to identify trends [13–15,20,21]. PICUP has done similar work with their organizers shortly after the development of the community [22]. Similarly, researchers have used social network analysis (SNA) to understand community structure and patterns in FLCs/FOLCs [23–25] as well as communities like computer-supported collaborative learning [26–28]. These articles utilize common metrics in SNA to describe the structure of the social interactions in these communities through identifying clusters of participants, pairs of participants, highly connected participants, less active participants, and more. SNA gives insight into the communication patterns of the community and allows us to study the structure of those patterns [29–32]. Dou and Zwolak [33] present a comprehensive introduction to SNA.

We use social network analysis in this work to study the patterns of communication between members of the PICUP community on Slack. While this community is less rigid than traditional CoPs, as we will describe in Sec. II, PICUP's participants use Slack to discuss ideas beyond in-person meetings and thus represent a unique subset of the PICUP community. Similar to the works mentioned, we will use SNA metrics, such as strength centrality, closeness centrality, reciprocity, and global clustering coefficient, to characterize the nature of communication in the Slack Workspace. We use these metrics to assess whether PICUP is meeting its defined goals (Sec. III). Through identifying if and how PICUP is meeting its goals in the confines of the Slack Workspace, we will be able to inform the PICUP community about changes they may want to make and motivate future analysis of the community.

We will answer the following research questions:

- What patterns of communication do we observe about the PICUP community and their Slack interactions through a social network analysis lens?
- In what ways do these patterns align or misalign with PICUP's main goals?
- What changes can be made to PICUP based on this analysis? What recommendations can be made to other FOLCs when using these platforms?

Through answering these questions, we will also clearly state the limitations of our study and identify other research methods that can be used to assess PICUP's goals, beyond Slack.

The rest of the paper is structured as follows. In Sec. II, we provide additional context on the PICUP community and the associated Slack data. We describe the steps taken to restructure the data for network analysis and show examples for converting this data to a network structure. In Sec. III, we will outline our conceptual framework by first defining various SNA metrics and tying them to PICUP's goals. In Sec. IV, we construct networks and calculate the defined metrics for various channels in the PICUP Slack Workspace. Section V describes the results of our analysis and the implications for the leaders of PICUP, and Sec. VI states the limitations in our analysis and related claims.

II. DATA CONTEXT AND PREPARATION

In this section, we provide additional context on the structure of the PICUP community, particularly their use of Slack. We describe our data preparation process of converting the raw Slack messages into a SNA friendly format. Then, we show the step-by-step process of representing the Slack messages as a social network while describing the various choices for representing communication in a network. Our detailed transparency in this process is to highlight what information is captured by these networks (and subsequently, what is not). These network representations can show a condensed visualization of the communication in the Slack channel but are not the sole mechanism for extracting information.

A. PICUP community structure

The creators of PICUP developed the group as a space for instructors to learn and communicate with one another about pedagogical tools for integrating computation into physics curricula. In the early days of PICUP, Irving and Caballero [22] demonstrated the community had potential to meet the framework of the traditional CoPs and FOLCs. While PICUP meets many of the definitions for CoPs and FOLCs, it has a less rigid structure than traditional settings. PICUP offers workshops, meetings, and meetups at conferences throughout the year, but these are not required meetings for the group. The group is ongoing, and anyone can join at any time. There are resources shared with the community and workshops for helping educators implement them, but instructors can also pull directly from the site and implement them on their own. They are united by the common interest of integrating computing into their physics-focused classrooms, but there is a wide depth and breadth in the types of instructors that engage with PICUP's materials. PICUP offers a supportive community to engage and learn from experts, but this happens through formal means like the workshops as well as informal means like the Slack Workspace. Following the

traditional CoP framework, there are opportunities for instructors to move from the periphery to become more active, expert-like members of the community, but this is dependent on the engagement and interest of the individual and not inherent to the community.

With this in mind, PICUP creates a community of researchers with likeminded goals but does not fit into the traditional, rigid definitions of communities of practice and faculty online learning communities. This paper does not seek to define what PICUP currently is, evaluate further what components of FOLCs it meets, identify if subgroups of PICUP follow CoP, or if it has the potential to be classified as a traditional FOLC. The project has been funded by another grant, which will seek to answer some of these questions.

The Slack Workspace also does not follow a traditional CoP or FOLC framework. Rather, it is a component of PICUP and is a mechanism for further communication. While it may help enhance PICUP's FOLC status and aid its members in becoming expert level, this is not something we can evaluate directly without analysis of the entire PICUP community and support structure. Additionally, the Slack community only represents certain users and the experiences of those users in a particular environment. While SNA will give insight into communication patterns, they are restricted to the community interactions on Slack. Because PICUP's Slack community cannot be mapped directly to CoP, we have defined our own conceptual framework connecting the community, SNA metrics, and PICUP's goals (Sec. III). Our analysis in this work will additionally justify why PICUP does not meet the traditional definitions for FOLCs or COPs.

B. From Slack messages to processed data

While SNA representation does not retain the content of the Slack messages, we take several steps to ensure as much of the communication is retained as possible. In this subsection, we describe the process to restructure the Slack messages into a usable format for SNA.

Slack (Searchable Log of All Communication and Knowledge) is an instant messaging platform used by many research teams and businesses [34]. A workspace hosts a community for users in a specific team or company. Within the workspace, users can create channels for communication dedicated to certain topics. Channels can be made public or private. Last, users can message another or multiple users directly—outside the channels. A user can be a part of multiple workspaces and multiple channels within that workspace.

While the PICUP Slack Workspace is still open, our data contains messages from August 2016 to June 2021. The total number of messages sent in the entire workspace prior to any data reformatting is 9139. Our work only includes messages sent in public channels, not messages sent in private channels or direct messages. (We do not include messages sent in

private channels as we do not have access to that data.) The total number of users at the end of data collection was 475. When a user joins any Slack Workspace, they are automatically added to the “General” and “Random” channels. Users must opt-in or be invited to other channels in the network.

In Slack, a user may use tags to notify specific users in the channel. For example, a message containing “@user1” is visible to everyone in the channel, but only notifies user1 about the given message. Messages that tag a specific user only result in one entry to the dataset per user. For example, if user2 sends the message, “@user3 can you please send me the textbook you are using for your intro to physics class?,” our data would contain an entry only from user2 to user3 with the associated message content and time stamp.

Users can also send a message to all the other users in a channel using “@channel” in the given message. Each @channel message is translated into multiple entries between the sender and all current users in the channel. For example, if user1 sends a message, “Hi @channel, I have a question about Python!,” our dataset would contain a single entry from user1 to every current user in the channel with the same timestamp and same message. All messages that do not contain a tag to a specific person, even if they do not include an @channel tag, are treated in this manner. To avoid alerting the whole channel with a message in response to another message, a user can “reply in thread.” This creates a dialog between the sender and any other user who chooses to message in the thread. As we did before, we create a new entry between the sender and the current users in the thread with the message sent.

In our dataset, the last column “To_Channel” indicates whether the message was sent to the entire channel (true) or directed at a specific person (false). For messages tagged for a specific person, the “Userid_To” column contains the user who was tagged. For messages that were sent to the whole channel, the row value is repeated for the total number of users in the channel, with the “Userid_To” column individually filled with each individual user ID.

Additionally, any user may respond to a message with one or multiple emojis. Because the emoji reactions are still a form of communication between two users, it is crucial to retain this information in the network representation. Thus, we translate each emoji reaction into a message between the two users. For example, if a user reacted with a “:vibepartycat” emoji [35], we generate a message from the receiver to the sender with the emoji description as the content. Other smaller cleaning processes include adding the file name to the message where files were attached, adding communication type where possible (e.g., join/leaves), and removing bots from users and messages. Table I shows a snapshot of this cleaned data.

At the end of the data processing stage, we separate the data by channel to facilitate network building. There are 44 total channels in the Slack Workspace. Only 13 of these channels have more than 40 messages prior to data restructuring. In Table II, we select a few channels to

TABLE I. A snapshot of the restructured data from the Glowsript channel. The table below displays the post-processed data following our steps to identify receivers and senders, transform messages sent to the entire channel, and represent emoji responses as messages.

	Date	Time	Userid_From	Message	Userid_To	Channel	To_Channel
0	April 11, 2017	11:51	U1YKKMTFE	likes	U1YEMA4BA	glowsript	False
1	May 16, 2017	10:42	U1YKM7ZRT	likes	U1YKKMTFE	glowsript	False
2	May 16, 2017	10:42	U1YEMA4BA	likes	U1YKKMTFE	glowsript	False
3	April 11, 2017	11:51	U1YEMA4BA	Agreed, that doesn't work! If...	U1YKKMTFE	glowsript	False
4	April 11, 2017	11:53	U1YKKMTFE	Yes, I ended up using...	U1YEMA4BA	glowsript	False
5	April 12, 2017	11:48	U4GVBA31T	@U1YKKMTFE the ball object...	U1YKKMTFE	glowsript	False
6	April 12, 2017	13:15	U1YKKMTFE	If that is true, then what should...	U4GVBA31T	glowsript	False
8	April 12, 2017	13:21	U4GVBA31T	maybe they hadn't finished the...	U1YEMA4BA	glowsript	False
9	April 12, 2017	13:24	U1YKKMTFE	OK, good. In 2.3 I get...	U1YEMA4BA	glowsript	False
10	April 12, 2017	13:24	U1YKKMTFE	OK, good. In 2.3 I get...	U4GVBA31T	glowsript	False
11	April 12, 2017	13:36	U1YEMA4BA	I just went to glowsript.org...	U4GVBA31T	glowsript	False
12	April 12, 2017	13:36	U1YEMA4BA	I just went to glowsript.org...	U1YKKMTFE	glowsript	False
13	April 12, 2017	13:41	U1YKKMTFE	Oh, I was on Trinket	U1YEMA4BA	glowsript	False

TABLE II. The table shows the channel name, number of raw messages (i.e., distinct messages sent in Slack), number of messages post data restructuring (i.e., the total number of interactions), the total number of users, and the number of users that have sent or received at least one message. As seen, the number of users and the number of messages sent to the entire channel highly affect the number of messages in the post data restructuring column.

Channel name	Raw data	Total interactions	Total number users	Users with ≥ 1 message
Advanced Thermodynamics	60	678	30	16
Advanced Mechanics	49	770	34	24
Why We Do This	83	1224	35	23
Upper Mid-West	247	3582	31	30
Classroom Pedagogy	102	3338	72	58
Trinket	133	1839	38	37
Glowsript	476	17 618	92	91
Jupyter	776	28926	128	120
Random	966	163 296	459	457

compare the number of messages in the pre-processed and post-processed data to highlight the variance in activity between the channels. The channels are presented by increasing the total number of users in the Slack channel, with Advanced Thermodynamics being the smallest and Random being the largest. The associated networks for each of the channels are in Sec. IV and in Appendix A. For networks of similar size (in both number of preprocessed entries and number of users), we see similar trends in the number of postprocessed entries. For example, there is about a $12\times$ increase in the number of messages between pre- and postprocessing for the Advanced Thermodynamics and Trinket channels. These channels also have about the same number of users, 30 and 37, respectively.

The number of messages in the post-processed data is heavily dependent on the number of users in the channel and the number of messages sent to the entire channel. Each message sent to the entire channel contains an entry to every other user in the channel. Thus, for a large number of users and even a moderate number of messages sent to the

entire channel, the total number of messages increases dramatically. Comparing the Jupyter and Random channels, while they have a similar amount of messages prior to data processing, the Random channel has nearly $5.5\times$ more entries post data processing, due to the channel containing about $3.5\times$ more users. These types of messages contribute most to the drastic increase in the total number of messages compared to messages tagging a user directly or an emoji reaction.

We note that the total number of users in Table II might be different from the total number of users seen in the figures below. Slack generates a message each time a user joins and leaves the channel. While these messages are in our dataset, we do not use them to construct the networks as they cloud the main interactions between users. If users join at the end of data collection, and there are no additional messages following their join message, they will appear in our total count of number of users; however, they will not appear in our network representations as they have not been a part of any interactions in the network yet. The number of

users with at least one message in the network can be seen in the last column of the table.

C. From processed data to networks

In Fig. 1, we illustrate how networks are generated from postprocessed data. The table is an example postprocessed data, and the right is the associated network representation. We are using *directed and weighted* networks. Each user is represented by a node and a corresponding number. Messages are represented with directed edges to indicate the sender and receiver. For example, message A is represented with an edge from user 1 pointing to user 2.

If multiple messages are sent between the same two users, we increase the weight of the edge. Message D is an example of this. Notice, the solid edge from user 1 to user 2 is thicker than the solid edge from user 3 to user 1 in message C. We use edge type to denote the type of message. Each of these first five messages was directed at a specific user, and therefore, the associated edge is shown as a solid line. The remaining rows in the table represent a message sent from user 1 to every other user in the channel—represented with a dashed edge.

The edge color indicates whether there is a reciprocated edge between two users. In our example, User 1 has sent message B and message E to user 3, and user 3 has sent message C to user 1. Because both users have sent a message to the other, the edge is colored gray. Opposingly, user 2 has not sent a message back to user 1, and hence, the edge is colored black. As we will see in the more complicated networks, this simple coloring can help us contextualize the network quickly.

The node size and node color are used to represent node strength. We present a formal definition in Sec. III A. Put simply, strength is the sum of the weights of the edges associated with a specific node. The in-strength of a node is the sum of the weights of the edges pointing toward the node, and the out-strength is the sum of the weights of the edges leaving the node.

From	Message	To	To Channel
User 1	Message A	User 2	False
User 1	Message B	User 3	False
User 3	Message C	User1	False
User 1	Message D	User 2	False
User 1	Message E	User 3	False
User 1	Message F	User 2	True
User 1	Message F	User 3	True
User 1	Message F	User 4	True

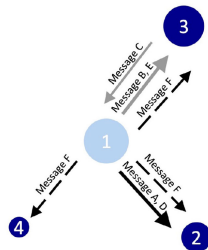


FIG. 1. We present an example to show how the structured data is represented as a network in our project. Users are represented with nodes, and messages are represented with directed edges, using size and color to indicate user attributes. In our representation, solid edges indicate direct messages between two users. Opposingly, dashed edges indicate a message that was sent from one user to the entire channel. Edge width corresponds to the number of messages between two users.

While we visually separate the edges sent directly to another user and to the entire channel, they are combined in the strength calculation. In Fig. 1, Node 3 has a total strength of 4 with an in-strength of 3 and an out-strength of 1. Nodes with a larger out-strength than in-strength (sending more messages out than receiving) are colored light blue, such as node 1. Users with a larger in-strength are colored dark blue, such as nodes 2, 3, and 4.

III. CONCEPTUAL FRAMEWORK

In this section, we define our conceptual framework for mapping PICUP's goals to social network analysis metrics. To define this framework, we need to introduce and define the SNA metrics we use in this work. Then, we will present our conceptual framework by connecting PICUP's goals with the associated expectations in the network structure and the metrics we will use to assess these goals.

A. Social network analysis metrics

Network science has a variety of metrics to characterize the nature of communication in the network [29–31]. Most metrics were initially constructed for an unweighted and undirected network. However, with weighted and directed networks, it is critical to incorporate these aspects into the metric calculation as it can affect the value, and thus the interpretation, of the metric [30,33]. Often, there are many options for a single metric that account for directionality and weighted edges of a network, which allows researchers to pick a variation that fits what these features represent in their networks [36–40]. We will discuss the ones we have chosen for our networks in this section. All analysis was conducted in R [41] with specific packages cited below.

1. Strength

As described in Sec. II C, the strength of a node, i , is the sum of the weights of edges associated with that node

$$s_i = \sum_{i \neq j} w_{ij},$$

where w_{ij} is the edge weight between nodes i and j [29,30,42].

The in-strength is

$$s_i^{\text{in}} = \sum_{i \neq j} w_{ij}^{\text{in}},$$

where w_{ij}^{in} is the weight of the directed edge to node i from node j , and the out-strength is

$$s_i^{\text{out}} = \sum_{i \neq j} w_{ij}^{\text{out}},$$

where w_{ij}^{out} is the weight of the directed edge from node i to node j [29,42]. We use IGRAPH's built-in functions for these calculations [43,44].

2. Closeness centrality

While the strength metric identifies frequent senders and receivers of Slack messages, we calculate closeness centrality to identify the central users in the network [45,46]. In an undirected, unweighted network, a node's closeness centrality is the average of the shortest paths from the node to every other node [29,47].

To include edge weight, we must adjust the distance measurement between two nodes, and the simplest extension of the shortest path would be the smallest total distance between the nodes. However, this prioritizes paths with smaller edge weights, which is not always the correct assumption for the application. In some cases, like in our networks, a stronger edge weight between two users can represent a stronger connection, and thus, it should represent an easier path to traverse through. A user with multiple edges with large weights should have higher closeness centrality value than a user with multiple edges with low weights. See Opsahl *et al.* [47] for specific examples.

Following Opsahl *et al.* [47], we update the distance metric between nodes i and j for intermediary nodes h to be

$$d_{w\alpha}(i, j) = \min \left(\frac{1}{(w_{ih})^\alpha} + \dots + \frac{1}{(w_{hj})^\alpha} \right),$$

where we simply invert the edge weights between two nodes. Additionally, we incorporate a tuning parameter, $\alpha \in (0, 1)$, to balance the effect of the weights of the intermediary nodes. For smaller values of α , the longest distance is the path with the greatest number of intermediary nodes. For larger values of α , the number of intermediary nodes is not weighted as heavily compared to the strength of the edges.

With the inversion of the edge weights, the closeness centrality (CC) becomes

$$CC_{w\alpha}(i) = \left[\sum_j^N d_{w\alpha}(i, j) \right]^{-1},$$

and the interpretation as expected is preserved.

This metric can easily be extended to account for directed edges. For each node, we calculate a closeness centrality out (CC_out)—available paths only include edges leaving the node—and a closeness centrality in

(CC_in)—available paths only include edges coming toward the node. Users with a large CC_out are good broadcasters of information. They are central in the network and have short, heavily weighted paths to many other users in the network. Users with a larger CC_in are central in receiving information. Using the directed and weighted implementation, users can fall into one of the four categories displayed in Table III.

Similar to strength, these classifications will assess whether PICUP is meeting its goals by identifying active users in the network and their associated influence. We will likely see the users spread across the four categories, but ideally, in a well-connected network, many nodes would have a similar score.

We calculate the closeness centrality metric using Opsahl [48]'s “closeness_w” function.

3. Weighted reciprocity

In addition to highlighting reciprocal communication with gray edges in the networks, we calculate reciprocity to understand the total amount of reciprocal communication in our networks. Reciprocity is the probability that the opposing counterpart of a directed edge is also included in the edge set [31]. For undirected and unweighted networks, it is calculated as the number of reciprocated edges divided by the total number of edges [29]. This metric is bounded between 0 and 1, where networks with a larger reciprocity have a larger probability of reciprocated ties.

However, this implementation of reciprocity does not account for edge weight—just whether there exists a reciprocated edge between two users. To represent edge weight appropriately in the metric, we follow Squartini *et al.* [49] and define weighted reciprocity as

$$w_r = \frac{\vec{W}}{W} = \frac{\sum_i \sum_{j \neq i} \vec{w}_{ij}}{\sum_i \sum_{j \neq i} w_{ij}},$$

where

$$\vec{w}_{ij} = \min\{w_{ij}, w_{ji}\} = \vec{w}_{ji}$$

for pairs of nodes i, j .

More plainly, we obtain a single value for the denominator by summing all the weights in the network. For the

TABLE III. The four classifications of closeness centrality values. Generally, users with a larger out-closeness centrality have short outward paths to other users in the network and are thus good broadcasters of information. Opposingly, users with a large in-closeness centrality are frequent receivers of information.

CC value	Interpretation
Large CC_out, large CC_in	Good broadcaster, frequent message receiver
Large CC_out, small CC_in	Good broadcaster, few messages received
Small CC_out, large CC_in	Poor broadcaster, frequent message receiver
Small CC_in, small CC_out	Poor broadcaster, few messages received

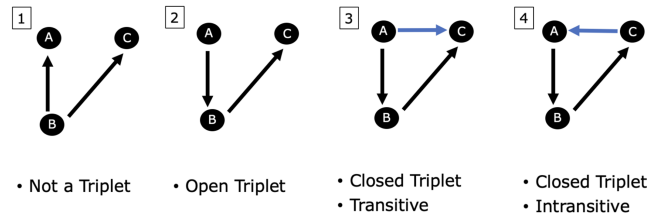


FIG. 2. Differentiating transitive and intransitive triplets. Triplets defined by Wasserman [32] are a 2-path—there is a path from node A to B and node B to C as seen in (2). A transitive triplet follows the transitive property implying that if there is a triplet from nodes A to B and B to C, the triplet is transitive if there exists a path from A to C [as seen in (3)] [36].

numerator, we generate a variation of the network to represent each interaction equally. If there is an edge from user 1 to user 2 with a weight of 3 and an edge from user 2 to user 1 with a weight of 5, the weight of both edges becomes 3. We sum these weights from this new network to obtain the numerator value. This weighted version is analogous to the original metric while accounting for the weights of the edges. This extension is still bounded between 0 and 1.

We use this metric to assess PICUP's goal of building and growing community in the Slack Workspace. High amounts of reciprocity, values closer to 1, indicate heavily reciprocated communication between the users and imply that there are relationships at least among pairs of users in the network. Ideally, to emulate a strong online community, we want users to be active and responsive to their peers. This metric does not identify the type of communication but gives insight into how much communication is happening in the network.

4. Weighted global clustering coefficient

Clustering coefficients are another common metric for characterizing the whole network structure. To identify areas of high and low density in a network, Watts and Strogatz [50] proposed the local clustering coefficient, which measures the likelihood the neighbors of a particular node are also connected. The local clustering coefficient of a node, v_i , in an undirected graph is the number of pairs of neighbors of v_i that are connected divided by the number of possible connections between pairs of neighbors of v_i , where neighborhood is the sub-network that contains all the degree 1 connections of v_i [29,32,36].

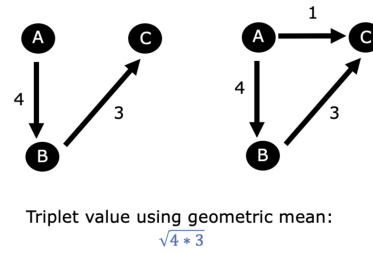


FIG. 3. For the weighted and directed GCC, we calculate a triplet value for each counted triplet in the metric. We use the geometric mean as shown. The geometric mean for an open and closed triplet with edge weights of 3 and 4 is $\sqrt{4 * 3}$.

To extend to a network-wide metric, we average the local clustering coefficient across all nodes in the network [38,50]. The global clustering coefficient (GCC) can be equivalently defined as

$$\frac{\text{Number of closed triplets}}{\text{Number of open triplets} + \text{Number of closed triplets}}, \quad (1)$$

where closed triplets are a set of three nodes and three edges with one connection between each of them, and open triplets are a set of three nodes and two edges [36,51]. See (2) and (3) in Fig. 2 for an example of an open and closed triplet.

By default, the global clustering coefficient above does not account for directed or weighted ties. Opsahl and Panzarasa [36] account for this in their version of the metric. For directed networks, our equation for GCC still follows Eq. (1); however, we change which triplets in the network get counted. In a directed network, an open triplet at node B occurs at (2) in Fig. 2—the directed edges must pass through B. Furthermore, a closed triplet can now be transitive or intransitive. As with a traditional definition of transitivity (if A is connected to B and B is connected to C, A must be connected to C), we call (3) in Fig. 2 a transitive, closed triplet. In this metric, we only count qualifying open triplets and closed, transitive triplets. To account for weighted edges, we continue to build off the previous definition by attaching a triplet value to each open and closed triplet. Equation (1) becomes

$$\text{WGCC} = \frac{\text{Total value of closed triplets}}{\text{Total value of open triplets} + \text{Total value of closed triplets}}. \quad (2)$$

There are many options for the triplet value; we use the geometric mean, as seen in Fig. 3. This metric is more robust against differences in unbalanced weights, and it bounded between 0 and 1.

We will use this value to assess PICUP's goal of continual community growth and strength in community building. WGCC values closer to 1 indicate many closed transitive triplets and thus indicate a high connectedness

among the community members. Ideally, in PICUP's networks, we want this value to be larger to indicate high amounts of communication between groups of three in the community.

We use Opsahl [48]'s "Clustering_w" function for this calculation.

B. Random networks and the configuration model

For the whole network metric calculations, we would like to study how (non)unique the value is compared to networks of similar size and distribution. This falls under the umbrella of inferential statistics, where we analyze the likelihood of observing a phenomenon about a sample from a population by chance.

In traditional hypothesis testing, we compare an observed result in the sample to some null value, where we are looking to identify if the sample result differs from what is assumed under the null hypothesis. However, the approximations needed to calculate these tests are only valid when the observations are independent—an assumption that does not generally hold true for social network data [52–54]. Specifically, we cannot generally assume that the individuals in the network do not influence each other's actions [55]. Additionally, while it would be ideal, we have no theoretical baseline distribution to compare our network metric value to. For example, we do not know the distribution of values for weighted reciprocity for networks with the same number of nodes and edges as our original network.

To statistically quantify our metric value, similar to a traditional z -test, we need to generate our own sample of similar networks to the reference (or original) network [53]. In this method, we are essentially estimating the probability distribution directly. Following the suggestion of [52,56], after generating an ensemble, we calculate the metric we are interested in for the original network and each of the networks in the ensemble. We then calculate the proportion of random networks in the ensemble with a lower metric value than the original network. Extreme probability values indicate that our network value is different than those in the sample (similar to a traditional confidence interval). Oppositely, probability values closer to 0.50 indicate our network metric is not different than networks of similar size, shape, and distribution. This value is an estimator of the likelihood that the metric value of the original network occurs by chance. With this method, we assume the generation technique creates a random enough ensemble such that the networks in the ensemble can be assumed to be independent, and thus, we can assume our probability value is reasonably accurate. For more details on this type of statistical analysis, please see Marion and Schreiber [57].

Often, we use a strength distribution to characterize a network [42,53,54,58]. As previously defined, a node's

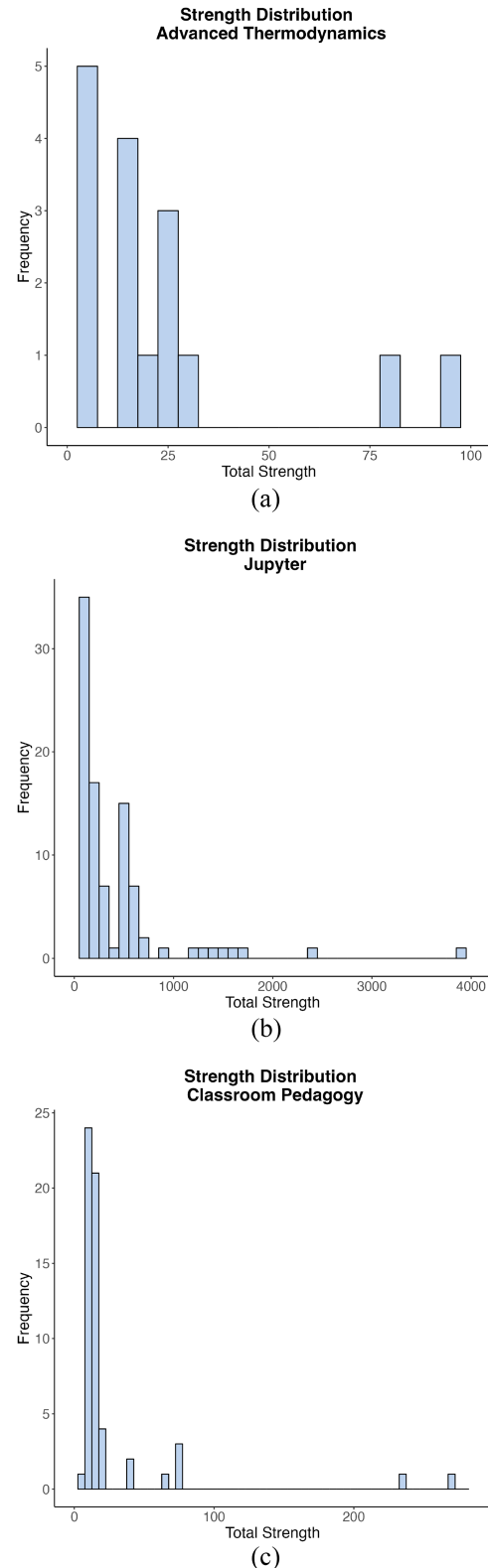


FIG. 4. The strength distributions for the main networks analyzed in Sec. IV—(a) for the Advanced Thermodynamics, (b) for the Jupyter, and (c) for the Classroom Pedagogy.

strength is the sum of the weights of the edges connected to the node. We use a histogram to represent the distribution of strength values for nodes in the network. Figure 4 shows the strength distributions for all three networks. We will perform an in-depth analysis in Sec. IV. The strength distributions are heavily right skewed. Often, these types of networks are called scale-free networks [31]. The strength distribution is a key choice in the method choice for generating new, similar networks, as it helps assess similarity between the original network and the networks in the ensemble. Additionally, the choice of network generation method can lead to false conclusions, especially with smaller amounts of data [53].

With this in mind, to create our random ensemble, we implement the configuration model [31,51,59]. The method preserves the strength distribution of the reference network as our distribution to generate random networks. In its simplest form, the configuration model can be thought of as a probabilistic reshuffling of the edges in the network. We fix the in- and out-strength for each node from the original network, as well as fix the number of edges. At each iteration, we pick a head and tail node with probability inversely proportional to the sum of the heads (tails). An edge is added, or the weight is updated between the head and tail nodes accordingly. For a walkthrough example with a smaller network, see Appendix B.

When implementing the configuration model, there is often a concern for repeated configurations in the generated ensemble. This is particularly true for small networks with few edges. Additionally, while the matchings of heads and tails occur with equal probability, the full network configurations do not occur with equal probability. To address this concern, we ran a simulation test to quantify the amount of repeated, identical networks in the ensemble of random networks. Due to the complexity of our networks, we are not concerned with repeated, identical networks. See Appendix C for more information.

For the Advanced Thermodynamics and Classroom Pedagogy channels, there are 7000 networks in our ensemble. For the Jupyter channel, there are 3000 networks in the ensemble. These numbers were chosen to balance computation time, as well as statistical accuracy, while minimizing the number of identical repeats.

C. An ideal network structure: Our conceptual framework

In this work, we are assessing three of PICUP's goals [7]:

1. Lowering barriers for faculty to integrate computation into courses.
2. Continual community growth and strength in community building.
3. Increased growth in the number of community leaders.

Ideally, continually meeting these goals will encourage more faculty to interact and engage with the community. In Table IV, we define our framework connecting PICUP's goals, the SNA metrics to assess these goals, and the metric values we expect if the goals are being met.

PICUP's main goal is to lower barriers for faculty to integrate computing into their classrooms. We expect the Slack community to aid in this endeavor by creating an accessible mechanism for communicating with others in the community. With low barriers, users in the network would be comfortable communicating with others in the network and sharing their ideas. To meet this goal, we would expect the networks to have many participants, many active participants, and many types of participants. This would imply that there is a lot of communication happening in the community, and the community is conducive to asking questions or sharing ideas. In the networks, we can assess this goal through metrics like strength, in-strength, and out-strength. Ideally, there would be a lot of users with large strength values and out-strength values greater than or comparable to in-strength. While ideal, this is not always expected for a traditional social network [60]. With that in mind, we would like this to happen with a large subset of the users in the network and have few users on the periphery of the network with large in-strength values.

Another goal of PICUP is to continue to grow the number of members engaging with the materials and create a strong sense of community. Again, the Slack should provide a remote opportunity for members to communicate and discuss ideas off-line and be an easy mechanism for engagement with others in the group. To meet this goal, we would expect high values for the whole network metrics—weighted reciprocity and weighted global clustering coefficient—particularly in relation to networks of similar

TABLE IV. The Conceptual Framework mapping three of PICUP's goals to expectations of community structure and metrics from social network analysis to assess whether the goals are being met in the Slack community.

PICUP goal	Expectation	SNA metric
Lowering barriers integrate computation into courses	Many active and types of participants	(High) strength, (Low) in-strength, (High) out-strength
Continual community growth	Reciprocated communication, many active users, many clusters	(High for all) weighted reciprocity, closeness centrality, strength, weighted global clustering coefficient
Increased number of community leaders	Active users beyond leadership	(High) closeness centrality, (High) strength across different levels

structure. High values would indicate much reciprocal communication and small clusters of users frequently communicating. Additionally, we would expect high values for closeness centrality and strength values. These would help identify the strength in the community structure in each of the channels and assess community growth.

PICUP is led by a group of experts in the computational physics field. However, building off the previous two goals, PICUP seeks to use its community to create more leaders and experts in their field. This is slightly more challenging for us to evaluate, as we are not conducting a temporal analysis of the networks. However, based on the previous metrics mentioned, we will be able to identify different types of users based on their participation and identify the current percentage of active users across different channels. The node-level metrics can help us identify the different categories of users we have across the networks. Since PICUP has been established for a fair number of years, we would expect the networks to have a fair number of users who are highly active to indicate there are leaders beyond the creators of PICUP, at least in the Slack community. Additionally, if PICUP was increasing the number of community leaders, we would expect many active broadcasters in the network who are sharing and facilitating information with other users.

In addition to the metric values, we can identify these features in the network representations. Ideally, we would see many users with a large node size and light blue color, indicating their active status in the network. Between these users, there would be gray edges with large weights to emphasize frequent, reciprocal communication. Oppositely, we would see few users with low strength on the periphery of the network. Ideally, we would expect a mix of messages sent to the entire channel and messages sent directly to another user. This balance would highlight a cohesive community in the Slack channel with users who feel comfortable chatting with each other, as well as the expert knowledge of specific users in the community.

In Sec. IV, we will assess whether PICUP has met the three identified goals on three networks using the metrics defined in Sec. III A. We conduct our analysis on three networks of varying size—both in number of users and messages—and varying channel topic focuses from pedagogy to computing. We will extend this analysis to five more channels to characterize the workspace as a whole.

As we will discuss more in Sec. VI, we are limited by the claims we can make on whether these goals have adequately been met, as we are only examining the Slack community, which is one aspect of the PICUP community. There are, of course, other types of analyses we can use to further assess PICUP's goals.

IV. ASSESSING PICUP GOALS USING SOCIAL NETWORK ANALYSIS

Before beginning our analysis, we note that the users of PICUP can join the Slack channel at any time, and thus, the

network is a snapshot of the communication in the channel up to that point. This framing influences how we analyze the networks. While a network may contain less active users on the periphery, this does not necessarily indicate that those users will always remain less active. They simply joined the channel shortly before we ended our data collection.

In the following sections, we analyze eight channels with a variety of topic focuses, number of users, and number of messages. We chose three of the eight to perform a deeper analysis on Advanced Thermodynamics, Jupyter, and Classroom Pedagogy. As we will explain in each section, the three channels represent three different network sizes (in terms of number of users and number of messages) and have three distinct discussion focuses.

A. Advanced Thermodynamics

The Advanced Thermodynamics channel is nearly the smallest channel in the Slack Workspace with 16 users, 60 raw Slack messages, and 678 messages post data restructuring. While the number of nodes and edges in this network is on the smaller side, we can still retrieve meaningful information from the analysis. The channel is pedagogically focused with users discussing resources, ideas, and strategies for integrating computing into their Advanced Thermodynamics curriculum. While users are talking about computing, the emphasis is on the integration of computing into the task rather than the computing tool directly. As we will see in each of the networks described, there are three clear subcategories representing three different levels of engagement in the network.

Extending the representation from Fig. 1, we represent the Advanced Thermodynamics channel with a directed, weighted network in Fig. 5(a). We use the rest of Fig. 5 to connect the elements of the Fig. 1 to the network representations. Figure 5(b) and legend objects (b)₁ and (b)₂ depict node size and color. As mentioned earlier, larger node sizes represent a larger strength value, and nodes with a larger out-strength than in-strength are light-blue in color. Figure 5(c) and legend objects (c)₃–(c)₅ represent nonreciprocal (black edges) and reciprocal (gray edges) communication, show edge weight (the number of messages between the set of nodes), and depict whether a message is sent to the entire channel (dashed edge) or to a user directly (solid edge). While this representation introduces complexity and requires additional explanation, it also allows for richer contextualization of the relationships within the Slack channel. We note all networks use the Kamada-Kawai layout in IGRAPH [61].

In the Advanced Thermodynamics channel, there are two light blue nodes. These highly active users in the network have the largest strength, the largest out-strength, and the largest CC_{out} (32.2, 27.7), and a large CC_{in} (5.18, 5.09). As expected, there are many edges—both gray and black—connected to these nodes. With their large strength and large closeness centrality values, these users—who are in

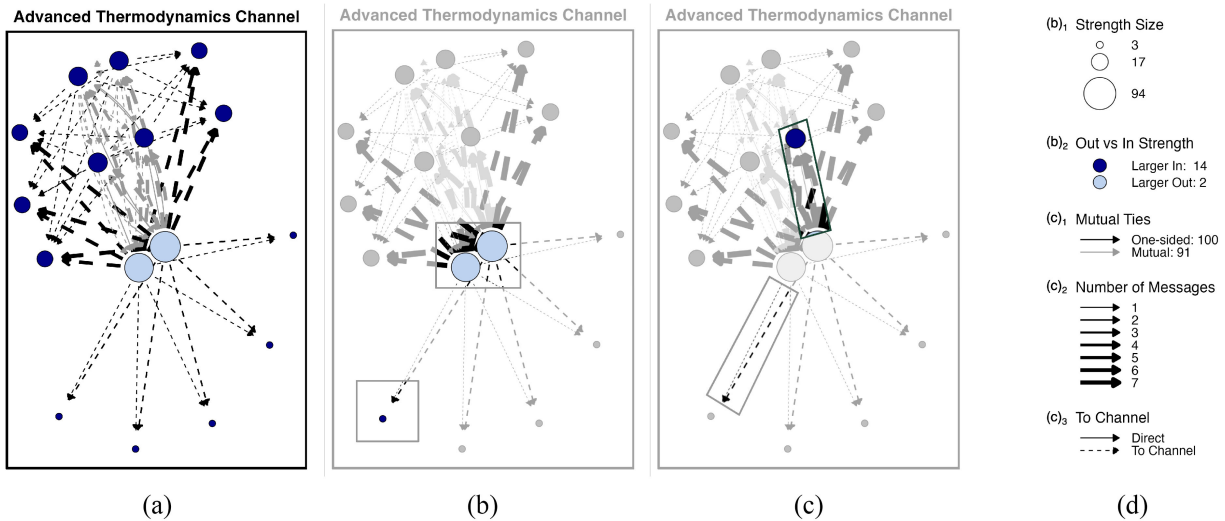


FIG. 5. (a) Shows the weighted, directed network for the Advanced Thermodynamics channel. Using the legend in (d), we highlight the representations for nodes (b) and edges (c) in the network. The three representations show the same network.

PICUP leadership—are in frequent communication with other users and are broadcasters of information.

The two highly active users are connected to four other users through many gray ties. These users have a larger in-strength than out-strength and have some of the largest CC_{in} values between 5.8 and 6.6, indicating they are frequent receivers of information. The six users connected by reciprocated edges signify a sub-community within the network.

Extending this part of the network to the five dark blue nodes with similar strength, we identify a third collection of users that have an out-strength of 0. These users have the highest CC_{in} values of 6.85. The users on the lower right of the network also fall into this category; they have joined the channel later, and thus, they have a smaller strength. While they have large CC_{in} values and have similar knowledge to other users in the channel, they are more passive members of the community.

Looking at the legend, despite only 2 users sending more messages than they are receiving, we have about the same amount of one-sided (100) and mutual messages (91). This is due to the sheer number of messages sent to the whole channel, with only about six users engaging in reciprocal conversations. This could be indicative of the users not using the tagging feature correctly, or truly wanting to share a message with the entire channel.

The weighted reciprocity value for the Advanced Thermodynamics network is 0.268. Looking at Fig. 6(a), this value is higher than the average value of the ensemble with a probability value of 0.996. This implies the reciprocity in our network is higher than other similar networks in the ensemble. However, with the weighted reciprocity metric bounded between 0 and 1, our value of 0.268 is on the lower end. Looking closely at the reciprocal

ties in the network, there are not many reciprocal relationships, and the reciprocal relationships that do appear are incredibly one-sided. In fact, most of the reciprocal ties between two users are highly imbalanced—an edge with a large weight for one user and an edge with a small weight for the reciprocal tie. Recall, the weighted reciprocity matrix is calculated by creating a symmetric network with minimum weight between the edges for two users, thus lowering the weighted reciprocity value.

From Fig. 6(b), the WGCC is 0.923. This value indicates the edges in the network that appear in closed triplets have a higher edge weight than those that do not. Furthermore, nearly all the triplets present in the network are closed [like (3) in Fig. 2]. This is fairly easy to see in the network. Most of the interactions from the middle to the top left are a part of at least one closed triplet. However, the users along the periphery in the bottom right are not connected to each other and likely represent the open triplets pushing the weighted GCC from 1. The high closed triplet value indicates intercommunication between clusters of three people in the network. However, most of the messages were sent to the entire channel. This makes it challenging to identify true clusters of users in the network, rather than those who are just a part of triplets due to sending messages to the entire channel. With this nature of messaging, the likelihood that a user is in a closed triplet is correlated to the time they joined the channel. This type of high value for clustering is typical in social networks [62].

Looking at the distribution of WGCC values for the ensemble created with the configuration model, the average WGCC is about 0.675, which is much smaller than the true value. With a probability value of 1, the WGCC for the Advanced Thermodynamics network is larger than all other networks in the ensemble.

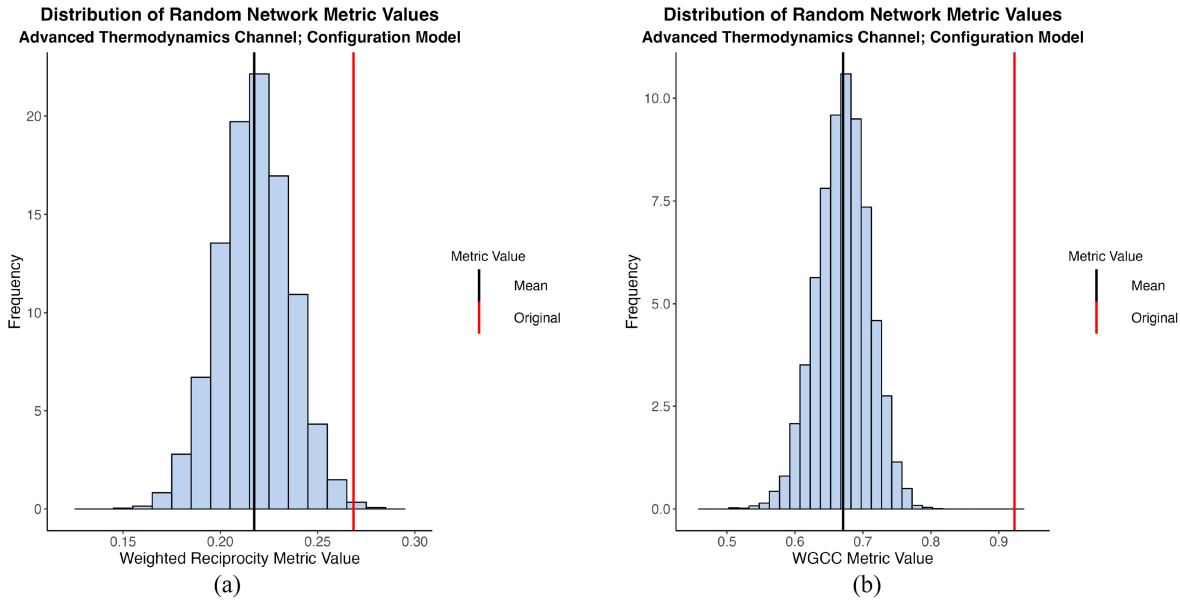


FIG. 6. (a) and (b) Applying the configuration model to Advanced Thermodynamics network with the weighted reciprocity and weighted global clustering coefficient metrics. The probability value for both is large indicating our original network has a much larger weighted reciprocity value and weighted global clustering coefficient compared to networks of similar size and distribution.

B. Jupyter

Besides “General” and “Random,” which by default include everyone in the Slack Workspace, “Jupyter” is the largest channel in the workspace with 119 users, 777 Slack messages, and 28 926 edges in the post-processed data. The channel is used to discuss questions and ideas related to Jupyter—an open-source software used for interactive computing with coding languages like Python, Julia, and R. Most commonly, it is used in Jupyter Notebooks. The interface intertwines coding blocks with written explanations in one file. With a low barrier to entry, they are a great pedagogical tool.

The Jupyter channel is a contrast to the Advanced Thermodynamics channel. Jupyter is one of the largest networks in the channel with $4\times$ as many users and $43\times$ as many edges as Advanced Thermodynamics. Additionally, the channel discussion is focused on a specific software tool rather than a specific course or field. While there is likely overlap in these channels as far as how to integrate this software into coursework, this channel is more heavily focused on usage of the software, including troubleshooting and error handling.

The weighted and directed network for the Jupyter channel can be seen in Fig. 7. While it is challenging to see minute details with the large number of edges, we can still extract the overall structure from the network. About 16% of the users have a larger out-strength, with about 35% of those users being within PICUP leadership. These users also have extremely large CC_{out} values. We notice few reciprocal edges in the top left portion of the network between the highly active users. Oppositely, 83% of the users on the periphery of the network in the channel are receiving more messages than they are sending. A subgroup of these users

has a nonzero out-strength and has some reciprocated communication with other users in the network, but nearly 50% of the users have an out-strength of 0. Furthermore, there are $1.7\times$ more nonreciprocated edges than reciprocated edges—many of which were sent to the entire channel.

In general, we are seeing a similar structure to the Advanced Thermodynamics channel. There are three types of users, with varying levels of engagement, and most of the communication in the network is passive and one-sided, as there are many black edges.

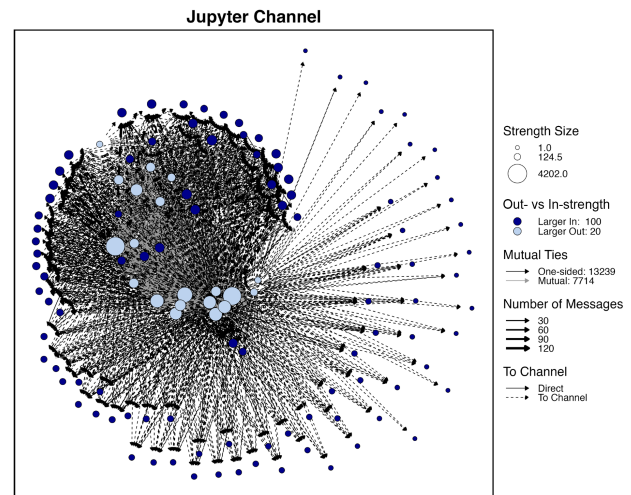


FIG. 7. We show the directed, weighted network representation of the Jupyter channel. Despite the drastic increase in size from the Advanced Thermodynamics network, we see similar network characteristics with three levels of users engagement.

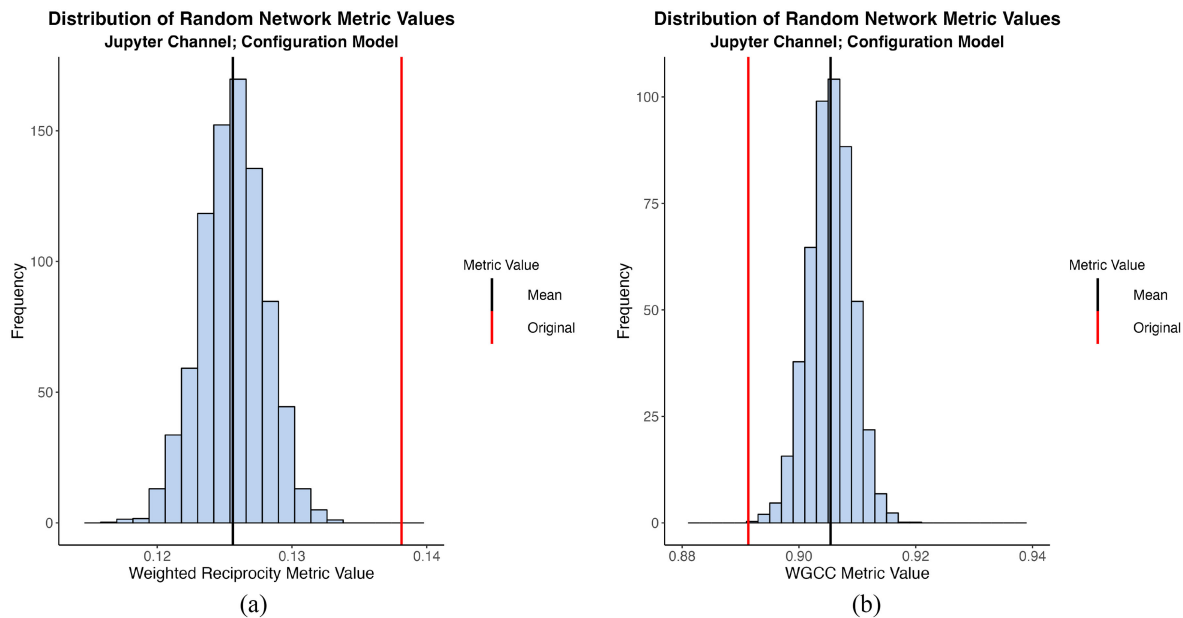


FIG. 8. (a) and (b) We apply the configuration model to the Jupyter network, calculating the weighted reciprocity and WGCC metrics. For weighted reciprocity, the probability value is 1, indicating more reciprocal communication than in the ensemble. For WGCC, the probability value is 0, indicating the value is smaller than other networks in the ensemble.

For the whole network metrics, we see similar values to the Advanced Thermodynamics networks. Looking at Fig. 8(a), the weighted reciprocity of 0.138 is higher than all other networks in the ensemble, with probability 1 at a value. While the value is large comparatively, it is low, highlighting the lack of a reciprocity network, particularly with the edges for the many nodes on the periphery. Additionally, the WGCC is 0.891, which is relatively high, but still lower than the networks in the ensemble.

C. Classroom Pedagogy

Finally, Classroom Pedagogy is a medium-sized channel with 72 users, 102 raw Slack messages, and 3338 messages post data restructuring. The channel provides a balance in the number of users and messages between the previous two channels and is also focused solely on pedagogical practices. While channels like Trinket and Glowscript may be a better network size comparison between Advanced Thermodynamics and Jupyter, their channel topics are too closely aligned with Jupyter.

In the Classroom Pedagogy network, Fig. 9, we see a similar structure to the previous two networks in relation to the types of users. There is a small portion of users who are highly active with large out-strength and large CC_out values, a larger portion of users with a larger in-strength than their nonzero out-strength, and an even larger portion of users with a zero out-strength. Despite the similar categorization of users, the proportion of reciprocated edges to nonreciprocated edges is much smaller—about 10%. This is also visible through the network which has a highly radial structure with many black edges.

In line with the previous two networks, the weighted reciprocity and the WGCC for the Classroom Pedagogy channel are 0.068 and 0.843, respectively as seen in Fig. 10. With probability 1, both are greater than similar networks in the ensemble generated by the configuration model. Again, the weighted reciprocity value is low, emphasizing the lack of reciprocal communication, as seen in the network. The WGCC value is lower than the previous two networks but still relatively high, indicating a fair amount of closed triplets in the network.

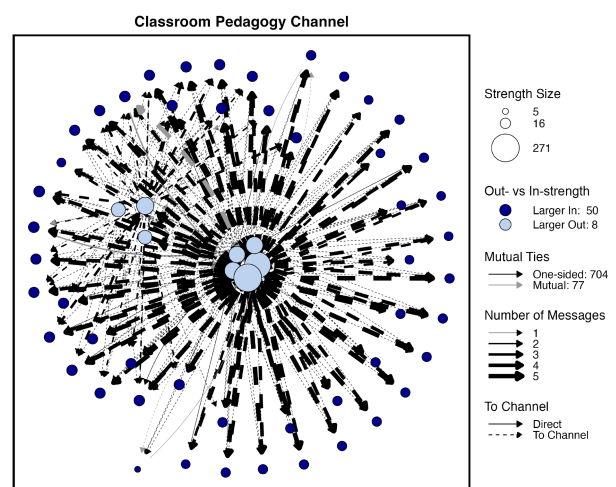


FIG. 9. The Classroom Pedagogy channel as a network. We see similar trends to the other networks in this paper. There are few highly active users and many users on the periphery.

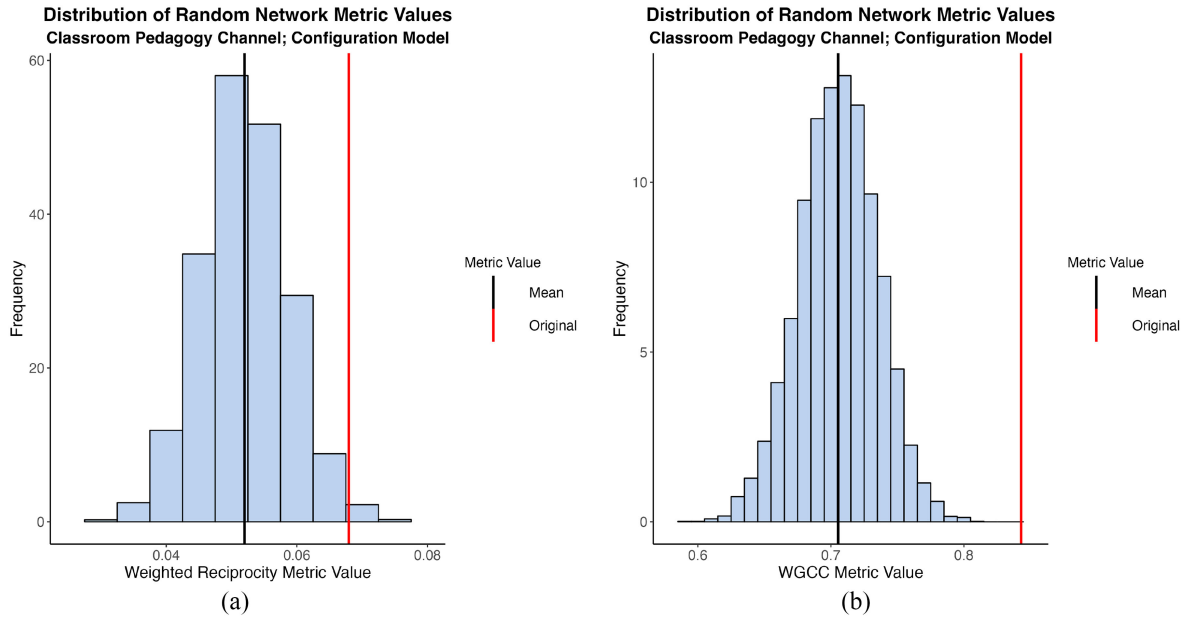


FIG. 10. (a) and (b) We apply the configuration model to Classroom Pedagogy channel. For weighted reciprocity, the probability value is 0.98, and for WGCC, the probability value is 1. Both values indicate the metric values are larger than other similar networks in the ensemble.

D. Other channels

In each of the three networks in Secs. IV A–IV C, we saw three different types of engagement:

1. Active users—Nodes with a larger out-strength than in-strength.
2. Passive users—Nodes with a larger in-strength than out-strength, and a non-negative out-strength.
3. Receivers—Nodes with nonzero in-strength, and an out-strength of 0.

Using these categories, we calculated the percentages of users in each group across eight channels in the Slack Workspace, seen in Table V. While not an in-depth analysis, we can explore trends of participation across multiple channels in the PICUP Slack. The receivers category is consistently the largest percentage of users (or equivalent to passive users in the Upper Mid-West channel). Often, the percentage of users in the first group is smaller than the second, with Advanced Thermodynamics, Trinket, and Jupyter as the exception. However, even in those channels, the average difference between the first and second categories is 8.3%.

Looking at these categories on a user level, we calculated the number of users in the same category across multiple channels. We found that only two users appear in the first category of more than 50% of the eight channels and 20 appear in more than 25% of the eight channels. Similarly, 10 users appear in the third category of more than 50% of the eight channels and 69 appear in more than 25% of the eight channels. While there is more consistency in the users that remain in the third category, there are not many users who are using multiple channels. Additionally, only 12 of the 174 (or 7%) total users in the eight channels are in more than 50% of the eight channels, with 8 users only in 62.5% of them.

V. IMPLICATIONS

Extending from Sec. IV, we connect our findings from our social network analysis to our conceptual framework to assess if PICUP is meeting their defined goals in the context of the Slack Workspace.

The first PICUP goal we identified is *lowering barriers for faculty to integrate computation into courses*. In the

TABLE V. Percentages for users in each of the three defined categories across eight channels in the PICUP Slack Workspace. The third category nearly always has the largest percentage of users, and the first category is the smallest in all but three channels.

Channel name	Advanced Mechanics (%)	Advanced Thermodynamics	Upper Mid-West	Why We Do This	Classroom Pedagogy	Trinket	Glowscript	Jupyter
Active users	16.7	12.5	26.7	17.4	12.5	24.3	16.5	16.7
Passive users	4.2	31.3	36.7	26.1	31.3	13.5	23.1	15.0
Receivers	79.2	56.3	36.7	56.5	56.3	62.2	60.4	68.3

Slack, low barriers would indicate users are frequently interacting with one another to ask questions and share ideas. To align with this, we expect the network to have a fair amount of active users, particularly with high strength values. Across all networks we analyzed, the group of active users was often the smallest percentage of the three categories (or at least far smaller than the receivers' category). This corresponds to our visual assessment few light blue nodes. Often, these users seemed to be driving the communication in the channel, with many of the reciprocal edges tied to them. This small group of users likely has a sense of community among themselves, but this clearly does not radiate through the rest of the channel.

The second PICUP goal is *continual community growth and strength in community building*. Similar to the first goal, we expect many active users in the network with high strength values. As mentioned, only a small percentage of users in each network are in the active users' category. These few users have a high out-strength and a high closeness centrality. The users send and receive many messages while also being good broadcasters of information to the rest of the channel. These users are essential to sharing information with the whole network and initiating conversations. While we do not expect all users to take on this role, having more users in this category would increase the amount of communication between users and strengthen the community.

In addition, to better assess the community in the network, we expect high values for reciprocity and the global clustering coefficient to represent frequent communication across all users in the network. Across all channels in Secs. IV A–IV C, the weighted reciprocity value was larger than the networks in the generated ensemble, but it was consistently low. With few active users and even fewer passive users, the low weighted reciprocity is not entirely surprising. As indicated by a few gray edges, there is little reciprocated communication in the networks, usually only between a small percentage of the users. The most active users in the network are often sharing messages with the community and receiving little-to-no responses. Opposingly, the weighted global clustering coefficient was fairly high for all three networks in Secs. IV A–IV C. While initially surprising given the other metrics, it is likely this metric is high due to the cluster of active users dominating the closed triplet total value as drivers of the conversation.

The final PICUP goal we assess is *increased growth in the number of community leaders*. The PICUP community has now existed for many years, and while the core creators continue to be involved, the community is meant to expand by creating more local leaders, which will further increase PICUP's reach. The mechanism for creating these leaders is often through in-person workshops; however, because the Slack community is an important aspect of PICUP, we would expect some of this leadership to translate to this domain. While we are not performing a temporal analysis, we would still expect to see this leadership reflected in both active users

and passive users, given the longevity of PICUP. Similar to the other two goals, we expect users to have high strength and high closeness centrality values. As explained, there are few users in each network that fit this description.

Combining these findings, we argue that PICUP is currently not meeting the three defined goals in the context of the Slack Workspace. For our three representative networks in Secs. IV A–IV C, we consistently observed low reciprocity, few active users, few broadcasters, many receivers, and many messages sent to the whole channel. In conjunction with our percentage calculations, the majority of users were receivers in all eight categories. This domination of users not interacting with other users in the network through messages or reactions highlights the lack of a whole network community and the number of community leaders. Furthermore, it does not appear that the PICUP community is engaging with one type of channel more than the other. This lack of interaction is consistent across channels focused on coding software *and* specific curriculum. This extends across multiple networks as well. As we found, only 7% of the total users in the Slack are in more than 50% of the channels. Users are self-selecting, which channels they would like to interact with, and then choosing not to engage through messages or reactions in those channels.

Together, these findings indicate a lack of community between the members of PICUP in the Slack Workspace. With few active users and many receivers, members of PICUP are not engaging in conversation with other members. However, this analysis does not allow us to understand and why this may be the case and what intentions the PICUP creators and users have for the Slack community.

VI. LIMITATIONS AND FUTURE WORK

As hinted at so far, while PICUP is not yet meeting its goals through our use of SNA to analyze the Slack community, there are limitations with this work and making broader claims. In this work, we only looked at the PICUP community interactions through Slack. As mentioned in Sec. I, the PICUP community has many mechanisms for interaction through both in-person and virtual means, beyond the Slack community. By limiting the analysis solely to the Slack interactions, we are limiting the interpretations we can make with regard to PICUP meeting its goals. Moreover, the use of Slack itself might be a limitation. While it is a popular platform for online communities, it may not be the best fit for PICUP's goals. The nature of Slack encourages quick, informal communication, which may not foster the deep, meaningful interactions that PICUP aims for. This could explain the high number of messages sent to the entire channel without much engagement from other users.

PICUP is a complex and unique community. As described in Sec. III, it does not fit squarely into traditional CoP and FOLC frameworks. The lack of a rigid structure allows for more fluidity in engagement from users but

makes it more challenging to critically assess how each component of the community aligns with PICUP's broader goals. The flexible boundaries for users hinder the ability to define who is a part of the community, in what ways they interact with the community, and what their goals are for being in the community. Additionally, while PICUP's broader goals are clearly defined, their specific goals for the Slack community are not. In this work, we constructed our own framework (in Sec. III) to use social network analysis techniques to evaluate how PICUP's overarching goals relate to the Slack Workspace. We found the engagement on Slack did not meet PICUP's overall goals. However, PICUP's goals for the Slack community might differ or be more nuanced than their overarching goals, which could affect our assessment.

PICUP might consider better defining its intentions for the Slack Workspace and how it is integrated into its community. Through qualitative analyses, researchers could interview PICUP's creators and current leaders to understand their intended purpose for creating the online community. Similarly, researchers should include responses from community members on their needs and usage of Slack, as they likely have different goals. Through clearer definitions, the PICUP community could more accurately evaluate whether they are meeting their goals and aligning their actions accordingly. It may be the case that Slack is simply meant to be a space for folks in the community to interact in an off-line format. However, if PICUP is looking to more closely translate the in-person experience to the off-line format, PICUP would need to make adjustments to how the Slack Workspace is managed to assist in creating a cohesive group and more community leaders.

In a similar vein, to accurately assess whether PICUP is meeting its goals in aspects like knowledge building and growth in leadership, we need to expand the data and type of analysis used in this work. There is an opportunity for the PICUP community to use analyses such as qualitative and/or natural language processing techniques to look at the content of the messages [63–66]. Conducting interview analyses with the current PICUP membership to gain insight into how they interact and engage with the material and the support that PICUP offers. These types of analyses, in conjunction with the SNA work presented, will give PICUP a comprehensive assessment of its community. Not seeing a trend in the Slack community does not imply that PICUP is not meeting those goals through other mechanisms.

Within the data specifically, messages with text and emoji reactions are treated equally in the dataset. The internet and social media age has altered how we communicate and share information with one another. An emoji response represents an acknowledgment or reaction to a message and, thus, still represents a form of communication between those two users. While this type of engagement might be different than a traditional, written response, denoting this in the data would have created additional

complexity in the network representations. Additionally, we chose to look at the network at a snapshot in time rather than a time series progression, creating multiple networks at different points in time. This choice prohibited us from tracking the growth of the PICUP community and their interactions over time. It is worth noting that the time stamps in this type of analysis would need to be on the order of months or years to see drastic progressions in the our data, especially for smaller channels.

Furthermore, we primarily chose to analyze each channel *in silo*. In this work, we compared the interactions of the PICUP community between different types of channels—both in terms of size (both number of nodes and number of edges) and main topic for the network. We studied smaller, class-specific channels like Advanced Thermodynamics as well as larger, coding-focused channels like Jupyter. Interactions in one channel may influence relationships and interactions between specific users in other channels. To understand community interaction across networks, we calculated the percentage of users in each of the three engagement categories. This provided some insight into the participation of users throughout the channels; however, understanding the influence and interactions between specific users would require a much more in-depth node-level exploration [67,68]. Similarly, our inability to analyze private channels restricts our understanding of user engagement outside of public channels and how this may differ from their public engagement.

In the networks, we chose to visually represent direct communication (solid) and to channel communication (dashed) separately, as they are distinct forms of communication. Multiple direct messages between two users may indicate a strong relationship between those two users, while multiple direct messages to a single user from many others in the channel might indicate an individual who has expertise in the given channel topic. While we cannot discern specifics about the interactions from the visual, representing the data in this way allows us to identify potential users this might be happening with. Similarly, messages sent to the whole channel both indicate there is a cohesiveness and strong comradery among the community or indicate the community is fairly new with no identification of potential experts in the group. Again, it is not possible to conclude which is true based on the networks alone, but the distinction of the edge types starts these discussions. We note this difference in message content is not accounted for in the statistical analysis—all messages are treated the same in these calculations.

Moreover, the weight of the edge accounts for the number of each of these type of messages. However, the metric values, such as strength, combine the total number of messages and, thus, do not take the type of message into account. While the edge type highlights a different type of communication, it is unclear if the users in the community are doing this properly—for example, are they appropriately

tagging users they want answers from directly, or are they messaging the whole channel without tagging a specific person? After looking briefly at the messages, it appeared that the latter may be the case, and we decided to calculate the values for the different types of messages together.

Furthermore, we use gray edges to indicate reciprocal communication between the users. An edge is colored gray if there is a shared message between two users, either direct, to channel, or a combination of both. Qualitatively speaking, two private reciprocal messages may be different in meaning than two to channel reciprocal messages, but as mentioned, it is challenging to quantify this correctly if users are not properly tagging those they want to converse with.

In addition to limitations and potential future analyses, we end this section with recommendations to increase engagement in the Slack Workspace. There are two main techniques for facilitating communication—encouragement and guidance. Encouragement requires active users and leadership to encourage users to participate in activities related to the community [69,70]. In PICUP, this could look like PICUP leadership encouraging users to share how they implemented a certain tool in the classroom or encouraging them to answer others' questions. This type of encouragement can build the confidence of users and, hence, lead to increased use of the community. Through guidance, leadership, and active users can help newcomers assimilate into the community quickly by setting norms, guidelines, and expectations for communication [69,70]. As described by Elliott [71], asynchronous communication is only as good as the users in it and as long as the interest is maintained. We note there is a lack of literature on strategies for encouraging participation in FOLCs in an equitable and inclusive way. General recommendations for online community forums include inviting people from diverse communities and encouraging all users to introduce themselves [72].

VII. CONCLUSION

In this work, we used social network analysis to examine the communication between members of the PICUP Community on Slack and assess if their engagement aligns with PICUP's goals. We observed three levels of engagement for PICUP users: active users, passive users, and receivers, where receivers were the most prominent category across all networks analyzed. There were very few active users in each network with high strength and high CC values. These users dominated the conversation in each network through frequent reciprocal communication with other active users and broadcasting messages to the channel. However, across the whole network, there were generally low reciprocity values despite these values being larger than networks of similar size.

The patterns observed indicate PICUP is not meeting its goals of lowering barriers, community growth, and

increased community leaders in the context of the Slack Workspace. There are few active users, particularly users beyond leadership, as well as limited reciprocal communication. Further, PICUP users do not frequently appear in multiple channels and tend to stay in the same engagement level across channels. Through the metrics and visually inspecting the networks, the Slack Workspace does not facilitate sustained community engagement in an online format. An ideal network with a cohesive community network would include more active users and reciprocal communication within and across networks.

As mentioned in Sec. VI, the limitations of this research lend itself to future analysis. Our analysis of these goals is confined to Slack, and thus, future analyses such as text analysis and interview studies could complement this work and strengthen the assessment of PICUP's goals. Through a multi-method approach, researchers would understand the community's intentions and purpose for interacting with Slack, and thus, PICUP could adjust its actions and guidance of engaging the community accordingly.

We believe similar communities to PICUP can conduct an social network analysis to assess the cohesion and engagement in their community. For other communities, we encourage a clear definition of how the component being studied is situated in the context of larger community goals. Through our detailed study describing our choices for data preparation, network representation, conceptual framework, and network calculation, we anticipate this work serving as an example for other communities.

ACKNOWLEDGMENTS

Thanks to Daryl McPadden for their guidance on the early stages of this project. Thanks to Paul Irving and Savannah Wear for assistance with initial data collection and coding. This work was supported by NSF DUE-1431776, NSF DUE-1504786, NSF DUE-1524128.

DATA AVAILABILITY

In compliance with Institutional Review Board, the deidentified data and code related to this project can be viewed in this GitHub Repository [73].

APPENDIX A: NETWORK REPRESENTATIONS FOR OTHER CHANNELS

In Figs. 11–15, we show the network representations for Advanced Mechanics, Upper Mid-West, Why We Do This, Trinket, and Glowscrip channels, as described in Sec. IV D. We notice the coding focused channels (Trinket, Glowscrip, and previously discussed, Jupyter) have more interactions than other channels in the Slack Workspace.

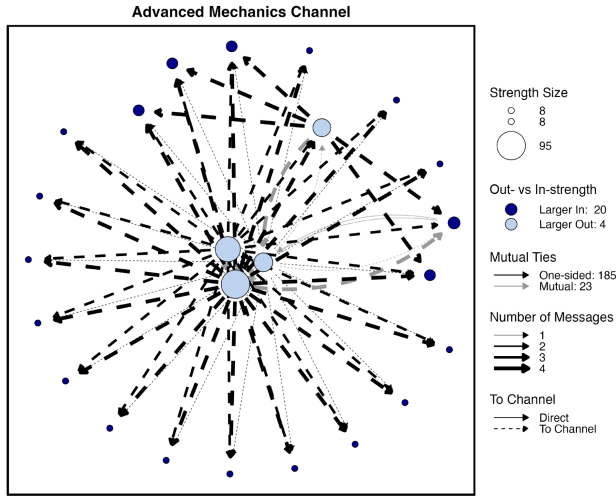


FIG. 11. Network representation for the Advanced Mechanics channel. The channel is similar in size to the Advanced Thermodynamics channel and similarly has a focus on curriculum of Advanced Mechanics.

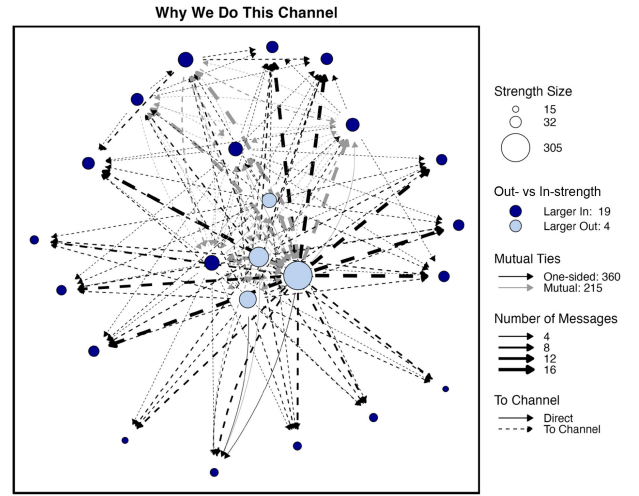


FIG. 13. Network representation for the Why We Do This channel. In the channel, users discuss their motivation for educating students and participating in this community.

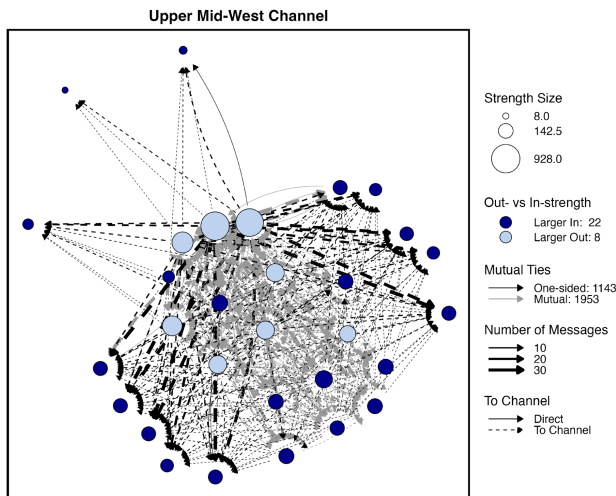


FIG. 12. Network representation for the Upper Mid-West channel. The channel is one of a few that are regionally focused.

APPENDIX B: CONFIGURATION MODEL WALKTHROUGH

In this section, we describe the steps of the configuration randomization method [31,51,59] through an example. The algorithm in pseudocode is presented in 1.

Figure 16(a) depicts an example network with four nodes (A, B, C, and D) and a set of edges connecting them. The table on the right contains the calculated values for the associated in-strength and out-strength of the nodes in the network. As mentioned, the configuration model preserves the strength distribution of the nodes in the network. For the sake of the generation method, we create a copy of the in- and out-strength lists, renaming that “heads” and “tails.”

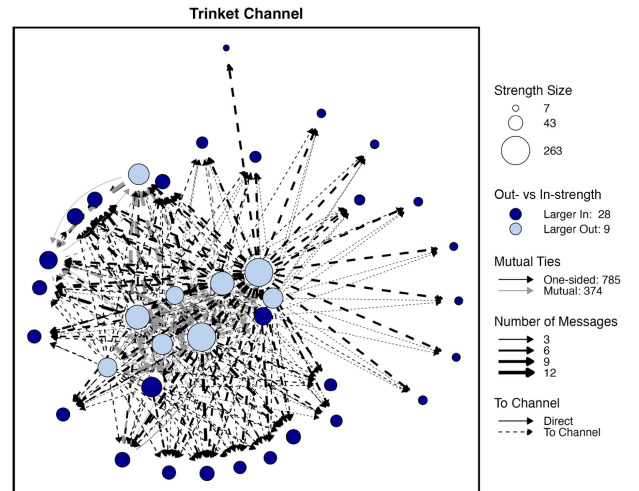


FIG. 14. Network representation for the Trinket channel. Trinket is another coding tool allowing users to run Python code interactively in a web browser.

As in Fig. 16(b), at the beginning of the algorithm, we start with the same node set as the original network. We then randomly select an index from the heads list and an index from the “tails” list. Here, node B is chosen as the head node, and node D is chosen as the tail node. Essentially, we are randomly picking two nodes to add an edge between, with the assumption that the in or out strength values are nonzero. We decrement the values associated with the chosen head and tail indices. Then, we add an edge from the chosen tail index to the chosen head index, creating a new edge if one does not already exist or simply incrementing the edge weight. This process is shown in Fig. 16(c), where we add an edge from node D to node B with edge weight 1.

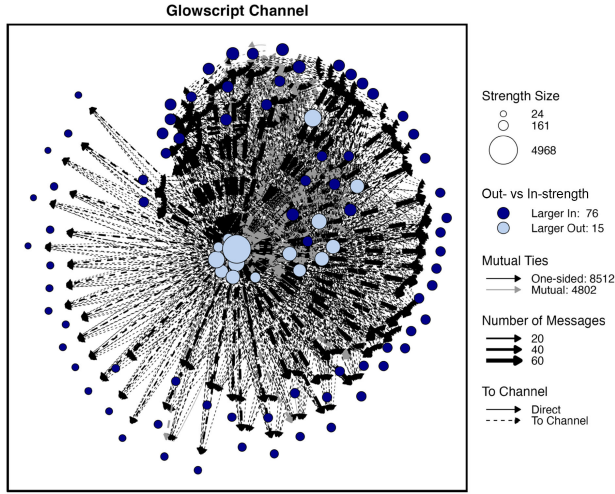


FIG. 15. Network representation for the Glowscript channel. Glowscript is a tool for creating visualizations in Python.

In Fig. 16(d), we show another iteration of the configuration model. In this iteration, node A is selected as the head node, and node D is selected as the tail node. In the same way, we decrement the values at those indices in the heads and tails lists and add the new edge.

We repeat this process of randomly selecting nodes to add an edge until there are no more available edges—meaning the sum of the in- or out-strengths list is 0. An example of a completed output from this randomization method can be seen in Fig. 16(e). As shown in the tables to the right, the heads and tails lists are zeroed out, the in-strength and

Algorithm 1. Configuration model.

```

1: heads  $\leftarrow [s_1^{in}, s_2^{in}, \dots, s_n^{in}]$ , where  $n$  = number of nodes
2: tails  $\leftarrow [s_1^{out}, s_2^{out}, \dots, s_n^{out}]$ , where  $n$  = number of nodes
3:
4: while sum(heads) > 0 do
5:
6:   head_index = random index s.t. heads[head_index]  $\neq$  0
7:   tail_index = random index s.t. tails[tail_index]  $\neq$  0
8:
9:   heads[head_index]  $\leftarrow$  heads[head_index] - 1
10:  tails[tail_index]  $\leftarrow$  tails[tail_index] - 1
11:
12:  if edge between head_index and tail_index exists then
13:    increase edge weight
14:  else
15:    add edge between head_index and tail_index
16:  end if
17: end while

```

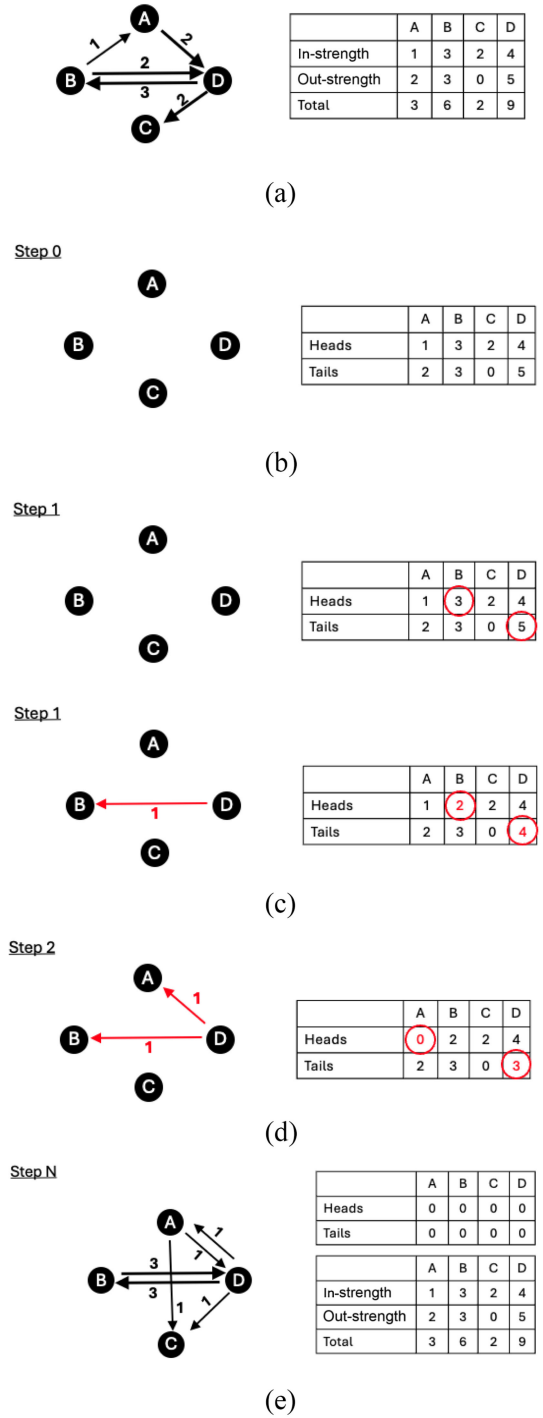


FIG. 16. A walkthrough of the first couple of steps of the configuration model on an example network. (a) Shows an example network with the associated in-strength and out-strength values. (b)–(d) Show the first couple of iterations of the randomization method. (e) Shows an example of randomized network preserving in and out strengths.

out-strength lists are identical to the original example network, and the network on the left has a different configuration than the example.

APPENDIX C: ADDRESSING REPEATS NETWORKS IN THE CONFIGURATION MODEL

As mentioned in Sec. III B, when implementing the configuration model, there is an upper limit on the number of possible network configurations. Repeated network configurations can affect the distribution of metrics values and thus skew the resulting probability value. To address the potential impact of repeated configurations on our analysis results, we conducted a simulation to quantify the number of identical networks in the random ensemble.

In the configuration model, the in or out strength associated with each node is fixed. Keeping this in mind, we only quantify the number of identical graphs, not particularly isomorphic graphs. We conducted three simulations: (i) a directed 4-cycle with edge weight 1, (ii) a directed 10-cycle with edge weight 1, and (iii) the Advanced Thermodynamics network.

For the directed 4-cycle, there are nine unique configurations preserving the in or out strength distributions, which is achieved with a small number of iterations (e.g., 25 iterations). Increasing to a 10-cycle, we generated ensembles of increasing sizes from 100 to 10,000. As seen in Table VI and Fig. 17, it takes an ensemble size of 2000 to see repeats of identical networks in the ensemble, where there is a repetition of 0.25%. Even for a large ensemble size of 10,000, the percentage of repeats is very low at 0.38%.

Generating an ensemble of 10,000 networks for the Advanced Thermodynamics network did not result in any identical repeats. Given that this network has more nodes, more edges, and higher weighted edges than the directed 10-cycle, this is not necessarily surprising. Thus, the likelihood of identical repeats in our ensemble for the Advanced Thermodynamics network—and subsequently, for our other networks that are larger in node and edge count—is very small. Thus, we are not overly concerned with the potential impact of identical repeats in our ensembles generated with the configuration model, and hence with the result that follows.

TABLE VI. The number of network configurations in the ensemble with the associated number of unique networks for a directed 10-cycle.

Total number in ensemble	Number of unique networks	Percentage of repeats (%)
100	100	0
250	250	0
500	500	0
750	750	0
1000	1000	0
2000	1995	0.25
3000	2996	0.13
4000	3992	0.20
5000	4992	0.16
7500	7489	0.15
10 000	9962	0.38

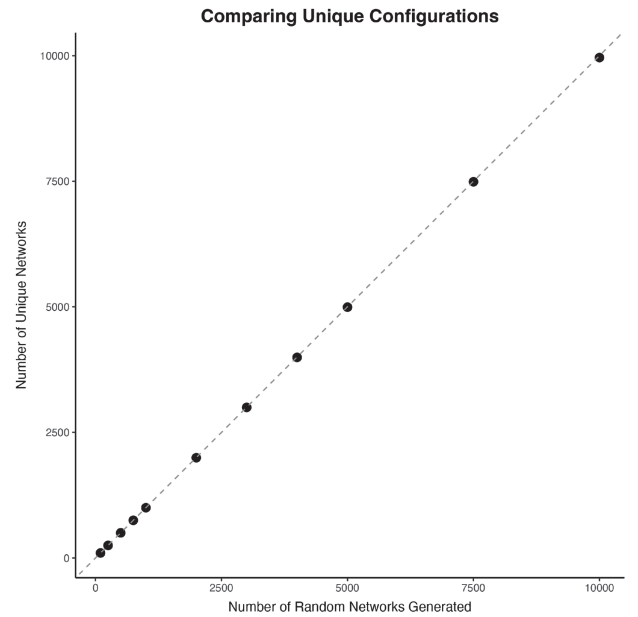


FIG. 17. Visualizing the number of networks in the ensemble with the number of unique configurations in the ensemble. The same information as Table VI. A nearly perfect identity line, with slight deviation at the higher values.

[1] D. Silvia, B. O'Shea, and B. Danielak, A learner-centered approach to teaching computational modeling, data analysis, and programming, in *Proceedings of the 19th International Conference on Computational Science-ICCS 2019, Faro, Portugal, 2019, Part V 19* (Springer, New York, 2019), pp. 374–388, [10.1007/978-3-030-22750-0_30](https://doi.org/10.1007/978-3-030-22750-0_30).

[2] A. J. Magana, M. L. Falk, and M. J. Reese, Jr., Introducing discipline-based computing in undergraduate engineering education, *ACM Trans. Comput. Educ.* **13**, 1 (2013).

[3] I. Lee, S. Grover, F. Martin, S. Pillai, and J. Malyn-Smith, Computational thinking from a disciplinary perspective: Integrating computational thinking in k-12 science,

- technology, engineering, and mathematics education, *J. Sci. Educ. Technol.* **29**, 1 (2020).
- [4] AAPT Undergraduate Curriculum Task Force Endorsed by AAPT Board of Directors, *AAPT Recommendations for Computational Physics in the Undergraduate Physics Curriculum* (American Association of Physics Teachers, College Park, MD, 2016).
 - [5] P. Heron and L. McNeil, *Phys21: Preparing Physics Students for 21st-Century Careers* (American Association of Physics Teachers, College Park, MD, 2016).
 - [6] N. Chonacky and D. Winch, Integrating computation into the undergraduate curriculum: A vision and guidelines for future developments, *Am. J. Phys.* **76**, 327 (2008).
 - [7] M. D. Caballero, N. Chonacky, L. Engelhardt, R. C. Hilborn, M. L. del Puerto, and K. R. Roos, PICUP: A community of teachers integrating computation into undergraduate physics courses, *Phys. Teach.* **57**, 397 (2019).
 - [8] A. Leary, P. W. Irving, and M. D. Caballero, The difficulties associated with integrating computation into undergraduate physics, [arXiv:1807.03581](https://arxiv.org/abs/1807.03581).
 - [9] <https://www.compadre.org/PICUP/>
 - [10] P. Baker, Creating learning communities: The unfinished agenda, in *The Social Works of Higher Education* (Pine Forge Press, Thousand Oaks, CA, 1999), p. 99.
 - [11] M. D. Cox, *Building Faculty Learning Communities: New Directions for Teaching and Learning: Volume 2004, Issue 97* (John Wiley & Sons, New York, 2004), Vol. 15, pp. 5–23.
 - [12] E. Wenger, *Communities of Practice: Learning, Meaning, and Identity* (Cambridge University Press, Cambridge, England, 1999).
 - [13] M. Dancy, A. C. Lau, A. Rundquist, and C. Henderson, Faculty online learning communities: A model for sustained teaching transformation, *Phys. Rev. Phys. Educ. Res.* **15**, 020147 (2019).
 - [14] T. L. Tinnell, P. A. Ralston, T. R. Tretter, and M. E. Mills, Sustaining pedagogical change via faculty learning community, *Int. J. STEM Educ.* **6**, 1 (2019).
 - [15] A. Furco and B. E. Moely, Using learning communities to build faculty support for pedagogical innovation: A multi-campus study, *J. Higher Educ.* **83**, 128 (2012).
 - [16] B. Iaquinto, R. Ison, and R. Faggian, Creating communities of practice: Scoping purposeful design, *J. Knowl. Manage.* **15**, 4 (2011).
 - [17] Y. Qian, S. Hambrusch, A. Yadav, and S. Gretter, Who needs what: Recommendations for designing effective online professional development for computer science teachers, *J. Res. Tech. Educ.* **50**, 164 (2013).
 - [18] J. A. Smith and S. A. Sivo, Predicting continued use of online teacher professional development and the influence of social presence and sociability, *Br. J. Educ. Technol.* **43**, 871 (2012).
 - [19] A. S. G. Erickson, P. M. Noonan, and Z. McCall, Effectiveness of online professional development for rural special educators, *Rural Spec. Educ. Q.* **31**, 22 (2012).
 - [20] C. J. De Leone, E. Price, M. S. Sabella, and A. G. Van Duzor, Developing and sustaining faculty-driven, curriculum-centered partnerships between two-year colleges and four-year institutions, *J. Coll. Sci. Teach.* **48**, 20 (2019).
 - [21] E. Price, A. C. Lau, F. Goldberg, C. Turpen, P. S. Smith, M. Dancy, and S. Robinson, Analyzing a faculty online learning community as a mechanism for supporting faculty implementation of a guided-inquiry curriculum, *Int. J. STEM Educ.* **8**, 1 (2021).
 - [22] P. Irving and M. Caballero, Understanding the PICUP community of practice, in *presented at PER Conf. 2017, Cincinnati, OH*, 10.1119/perc.2017.pr.042.
 - [23] C. Hatcher, E. Price, P. S. Smith, C. Turpen, and E. Brewster, Closeness in a physics faculty online learning community predicts impacts in self-efficacy and teaching, [arXiv:2209.09306](https://arxiv.org/abs/2209.09306).
 - [24] R. Cross, T. Laseter, A. Parker, and G. Velasquez, Using social network analysis to improve communities of practice, *Calif. Manage. Rev.* **49**, 32 (2006).
 - [25] S. Ma, G. L. Herman, M. West, J. Tomkin, and J. Mestre, Studying stem faculty communities of practice through social network analysis, *J. Higher Educ.* **90**, 773 (2019).
 - [26] M. De Laat, V. Lally, L. Lipponen, and R.-J. Simons, Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for social network analysis, *Int. J. Comput. Supported Collab. Learn.* **2**, 87 (2007).
 - [27] B. Wellman, Computer networks as social networks, *Science* **293**, 2031 (2001).
 - [28] X. Lin, X. Hu, Q. Hu, and Z. Liu, A social network analysis of teaching and research collaboration in a teachers' virtual learning community, *Br. J. Educ. Technol.* **47**, 302 (2016).
 - [29] S. P. Borgatti, M. G. Everett, and J. C. Johnson, *Analyzing Social Networks* (Sage, Thousand Oaks, CA, 2018).
 - [30] C. Kadushin, *Understanding Social Networks: Theories, Concepts, and Findings* (Oxford University Press, New York, 2012).
 - [31] M. Newman, *Networks: An Introduction* (Oxford University Press, New York, 2010).
 - [32] S. Wasserman, *Social Network Analysis: Methods and Applications* (The Press Syndicate of the University of Cambridge, Cambridge, United Kingdom, 1994).
 - [33] R. Dou and J. P. Zwolak, Practitioner's guide to social network analysis: Examining physics anxiety in an active-learning setting, *Phys. Rev. Phys. Educ. Res.* **15**, 020105 (2019).
 - [34] S. Butterfield, E. Costello, C. Henderson, and S. Mourachov, *Slack is Your Productivity Platform* (Slack Technologies, San Francisco, CA, 2009).
 - [35] <https://emoji.slack-edge.com/T02291UPX55/vibepartycat/2c7317181cfb7286.gif>
 - [36] T. Opsahl and P. Panzarasa, Clustering in weighted networks, *Soc. Networks* **31**, 155 (2009).
 - [37] T. Fardet and A. Levina, Weighted directed clustering: Interpretations and requirements for heterogeneous, inferred, and measured networks, *Phys. Rev. Res.* **3**, 043124 (2021).
 - [38] T. Schank and D. Wagner, Approximating clustering coefficient and transitivity, *J. Graph Algorithms Appl.* **9**, 265 (2005).
 - [39] R. Liu, S. Feng, R. Shi, and W. Guo, Weighted graph clustering for community detection of large social networks, *Proc. Comput. Sci.* **31**, 85 (2014).
 - [40] C. Wang, W. Tang, B. Sun, J. Fang, and Y. Wang, Review on community detection algorithms in social networks, in *Proceedings of the 2015 IEEE international conference on*

- progress in informatics and computing (PIC)* (IEEE, New York, 2015), pp. 551–555.
- [41] R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2021).
- [42] A. Fornito, A. Zalesky, and E. Bullmore, *Fundamentals of Brain Network Analysis* (Academic press, London, 2016).
- [43] G. Csardi and T. Nepusz, The IGRAPH software package for complex network research, *Inter J. Complex Syst.* **1.4.3**, 1695 (2006), <https://cran.r-project.org/web/packages/igraph/citation.html>.
- [44] G. Csárdi, T. Nepusz, V. Traag, S. Horvát, F. Zanini, D. Noom, and K. Müller, IGRAPH: Network analysis and visualization in R, R package version 2.1.4 (2025), [10.5281/zenodo.7682609](https://doi.org/10.5281/zenodo.7682609).
- [45] A. Bavelas, Communication patterns in task-oriented groups, *J. Acoust. Soc. Am.* **22**, 725 (1950).
- [46] L. C. Freeman, Centrality in social networks: Conceptual clarification, *Soc. Networks* **1**, 215 (1978).
- [47] T. Opsahl, F. Agneessens, and J. Skvoretz, Node centrality in weighted networks: Generalizing degree and shortest paths, *Soc. Networks* **32**, 245 (2010).
- [48] T. Opsahl, *Structure and Evolution of Weighted Networks* (University of London (Queen Mary College), London, 2009), pp. 104–122.
- [49] T. Squartini, F. Picciolo, F. Ruzzenenti, and D. Garlaschelli, Reciprocity of weighted networks, *Sci. Rep.* **3**, 2729 (2013).
- [50] D. J. Watts and S. H. Strogatz, Collective dynamics of “small-world” networks, *Nature (London)* **393**, 440 (1998).
- [51] M. E. Newman, The structure and function of complex networks, *SIAM Rev.* **45**, 167 (2003).
- [52] R. A. Hanneman and M. Riddle, *Introduction to Social Network Methods* (University of California, Riverside, CA, 2005).
- [53] E. A. Hobson, M. J. Silk, N. H. Fefferman, D. B. Larremore, P. Rombach, S. Shai, and N. Pinter-Wollman, A guide to choosing and implementing reference models for social network analysis, *Biol. Rev. Camb. Philos. Soc.* **96**, 2716 (2021).
- [54] D. P. Croft, J. R. Madden, D. W. Franks, and R. James, Hypothesis testing in animal social networks, *Trends Ecol. Evol.* **26**, 502 (2011).
- [55] S. J. Cranmer, P. Leifeld, S. D. McClurg, and M. Rolfe, Navigating the range of statistical tools for inferential network analysis, *Am. J. Political Sci.* **61**, 237 (2017).
- [56] S. Maslov, K. Sneppen, and A. Zaliznyak, Detection of topological patterns in complex networks: Correlation profile of the internet, *Physica (Amsterdam)* **333A**, 529 (2004).
- [57] R. Marion and C. Schreiber, Evaluating complex educational systems with quadratic assignment problem and exponential random graph model methods, in *Complex Dynamical Systems in Education: Concepts, Methods and Applications* (Springer, Cham, 2016), p. 177, [10.1007/978-3-319-27577-2](https://doi.org/10.1007/978-3-319-27577-2).
- [58] M. E. J. Newman, The physics of networks, *Phys. Today* **61**, No. 11, 33 (2008).
- [59] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Random graphs with arbitrary degree distributions and their applications, *Phys. Rev. E* **64**, 026118 (2001).
- [60] N. Sun, P. P.-L. Rau, and L. Ma, Understanding lurkers in online communities: A literature review, *Comput. Hum. Behav.* **38**, 110 (2014).
- [61] T. Kamada S. Kawai *et al.*, An algorithm for drawing general undirected graphs, *Inf. Proc. Lett.* **31**, 7 (1989).
- [62] M. E. J. Newman and J. Park, Why social networks are different from other types of networks, *Phys. Rev. E* **68**, 036122 (2003).
- [63] D. McHugh, S. Shaw, T. R. Moore, L. Z. Ye, P. Romero-Masters, and R. Halverson, Uncovering themes in personalized learning: Using natural language processing to analyze school interviews, *J. Res. Tech. Educ.* **52**, 391 (2020).
- [64] D. Wang, H. Wang, M. Yu, Z. Ashktorab, and M. Tan, Group chat ecology in enterprise instant messaging: How employees collaborate through multi-user chat channels on Slack, *Proc. ACM Hum.-Comput. Interact.* **6**, 1 (2022).
- [65] C. Costa Silva, M. Galster, and F. Gilson, A qualitative analysis of themes in instant messaging communication of software developers, *J. Syst. Software* **192**, 111397 (2022).
- [66] A. Martinez, Y. Dimitriadis, B. Rubia, E. Gómez, and P. De la Fuente, Combining qualitative evaluation and social network analysis for the study of classroom social interactions, *Comput. Educ.* **41**, 353 (2003).
- [67] S. S. Singh, S. Muhuri, S. Kumar, and J. Barua, From nodes to knowledge: Exploring social network analysis in education, *ACM Trans. Web* **19**, 1 (2025).
- [68] C. Wilson, B. Boe, A. Sala, K. P. Puttaswamy, and B. Y. Zhao, Beyond social graphs: User interactions in online social networks and their implications, *ACM Trans. Web* **6**, 1 (2012).
- [69] A. Tagarelli and R. Interdonato, *Mining Lurkers in Online Social Networks: Principles, Models, and Computational Methods* (Springer, New York, 2018).
- [70] J. Preece, B. Nonnecke, and D. Andrews, The top five reasons for lurking: Improving community experiences for everyone, *Comput. Hum. Behav.* **20**, 201 (2004).
- [71] J. C. Elliott, The evolution from traditional to online professional development: A review, *J. Digital Learn. Teach. Educ.* **33**, 114 (2017).
- [72] C. Abdullah, C. F. Karpowitz, and C. Raphael, Equity and inclusion in online community forums: An interview with Steven Clift, *J. Public Deliberation* **12**, 1 (2016).
- [73] egbolger, egbolger/picup_publication_codeshare: Initial release, 2025, https://github.com/egbolger/PICUP_Publication_Codeshare.git.