

RESEARCH

Open Access

Social network development in classrooms



Steven F. Wolf^{1*} , Timothy M. Sault¹, Tyme Suda² and Adrienne L. Traxler^{2*}

*Correspondence:

wolfs15@ecu.edu; adrienne.traxler@wright.edu

¹ Department of Physics, East Carolina University, 1000 E. 5th Street, Greenville, NC 27858, USA

² Department of Physics, Wright State University, 3640 Colonel Glenn Hwy., Dayton, OH 45435, USA

Abstract

Group work is often a critical component of how we ask students to interact while learning in active and interactive environments. A common-sense extension of this feature is the inclusion of group assessments. Moreover, one of the key scientific practices is the development of collaborative working relationships. As instructors, we should be cognizant of our classes' development in the social crucible of our classroom, along with their development of cognitive and/or problem solving skills. We analyze group exam network data from a two-class introductory physics sequence. In each class, on each of four exams, students took an individual version of the exam and then reworked the exam with classmates. Students recorded their collaborators, and these reports are used to build directed networks. We compare global network measures and node centrality distributions between exams in each semester and contrast these trends between semesters. The networks are partitioned using positional analysis, which blocks nodes by similarities in linking behavior, and by edge betweenness community detection, which groups densely connected nodes. By calculating the block structure for each exam and mapping over time, it is possible to see a stabilizing social structure in the two-class sequence. Comparing global and node-level measures suggests that the period from the first to second exam disrupts network structure, even when the blocks are relatively stable.

Keywords: Structural equivalence, Social networks, Network change

Introduction

Physics education researchers have been studying learning for quite some time to great effect (McDermott and Redish 1999). An early focus was on the cognitive domain of learning (e.g., Ginsburg and Oppen 1988). However, focusing on the cognitive aspects of learning does not provide us with the entire story. Indeed, many modern theories of learning posit that learning is also an inherently social process. These theories, such as social constructivism (Hirtle 1996), blend the cognitive domain with the social domain in order to better understand what is going on in our classrooms. While the environment that we use to teach—specifically, a classroom—has not changed significantly over most of the past millenium, our understanding of what goes on inside classrooms to foster meaningful learning has changed significantly—specifically, active and interactive classrooms promote deeper learning and more favorable learning outcomes (Hake 1998; Von Korff et al. 2016; Freeman et al. 2014; Mastascusa et al. 2011).

A common way that instructors promote active and interactive learning environments is by assigning and evaluating group work. Given the importance of the social domain in learning, some instructors have begun to engage in group assessment (Beatty 2015; Lin and Brookes 2013; Gilley and Clarkston 2014; Leight et al. 2012; Wieman et al. 2014). Therefore to understand social aspects of learning, we need to understand aspects of group work. However, in general, when studying group interactions during learning, researchers have focused on interactions within a single unit of people carrying out a task (e.g., Wolf et al. 2014). The issue with this approach is that it does not scale well with classroom size. There are simply only so many groups an instructor can meaningfully observe during a class period. Rather than limiting class size, one would hope to develop a set of tools for understanding and evaluating collaboration throughout a classroom. In this work, we begin to expand this study to the domain of the entire classroom community, utilizing the framework of social network analysis. We have observed collaboration patterns over the span of several group assessments and will describe how these networks change over time. We will probe these different classroom networks as they develop properties consistent with social networks such as the stratification of status and the development of roles.

This paper is organized as follows. We begin with a background of how networks have been applied to study education-related systems and, particularly, how networks have been used to describe group exam collaboration. We will then describe the classroom setting and how we collected these data. Next we discuss the methods that we have used for converting these data into networks and describe the network measures and methods that we have used to describe these networks and share those results. Finally, we summarize our results.

Background

Scholars in Physics Education Research (PER) have begun using networks to describe elements related to teaching and learning, most often focused on student interaction networks.

Networks in physics education research

Network analysis in higher education is not a unified body of research, but a set of occasionally overlapping subdomains with different methods and priorities (Biancani and McFarland 2013). These domains include faculty collaboration networks (coauthorship, citation, etc.), studies of college student success tied to network measures, and classroom-focused studies in specific disciplines. In PER, network analysis can broadly be divided into two categories: studies where networks are used to probe connections between ideas, and social network analyses where students in a course or courses are the nodes.

Networks of ideas

Though social network analyses of student collaborations are the most common use of networks in PER, a notable subset of studies cast more abstract entities as nodes. These have included concept mapping in domains such as electricity and magnetism to study coherence of knowledge structures (Koponen and Pehkonen 2010; Koponen and

Nousiainen 2018), or epistemic network analysis to study student explanations of solving computational problems (Bodin 2012). Some work compares the structures of problem organization networks between physics experts and novices (Wolf et al. 2012a, b). Other analyses have focused on answer co-occurrence networks on concept inventories (Brewer et al. 2016; Scott and Schumayer 2018; Wells et al. 2019). This kind of module analysis has also been applied to group exams (Sault et al. 2018). Finally, some work straddles social and conceptual networks, for example by following the flow of ideas in a conversation between students (Bruun 2016). These studies are all ways of accessing cognitive structures that organize ideas, either to test theories about those structures or to observe their development as students learn the material.

Student collaboration networks

The larger body of PER network studies, including this paper, treat students as nodes and interactions between them as edges. These interactions may be defined by their location, such as in a physics learning center (Brewer et al. 2012), outside of class (Zwolak et al. 2018), or on a discussion forum (Traxler et al. 2018). Other studies bound edge interactions by activity, such as collaboration networks for particular homework types (Vargas et al. 2018), or they may be based on broader prompts such as naming students “with whom you had a meaningful interaction in class during the past week” (Commeford et al. 2021).

The goals of these studies may be descriptive mapping but often are tied to student outcomes. A number of studies have found links between network centrality and students’ final grades (Traxler et al. 2018; Vargas et al. 2018), grades in a subsequent course (Bruun and Brewer 2013), or persistence in degree programs (Forsman et al. 2014; Zwolak et al. 2018). A smaller set of studies have focused on network structure. These might survey the development of student communities in the same class (Bruun and Bearden 2014; Traxler 2015) or look to compare features across different pedagogies (Traxler et al. 2020; Commeford et al. 2021; Brewer et al. 2010).

Previous exam network studies

Most closely related to this paper are studies performed on networks of students who have been collaborating in a group exam setting. These have examined both closed-collaboration (fixed groups set by the instructor) (Beatty 2015) and open-collaboration (students select their own groups) (Wolf et al. 2016, 2017) settings. In open collaboration settings, Wolf et al. found that the design of the room (e.g., movable desks in a tiered classroom or a flat classroom with large tables) changed the average size of groups (Wolf et al. 2016). In subsequent work, Wolf et al. studied the grade differential of all dyads, finding that, in classroom networks, grade is a proxy for status—at least at the end of the semester (Wolf et al. 2017). For each semester in the study sample, on the first exam, student grade differential distributions were not significantly different regardless of dyad type. However, on the final exam, grade differentials were significantly more positive for students in asymmetric dyads than they were for students in mutual dyads (Wolf et al. 2017). In other words, students with higher grades than a dyad partner were less likely to reciprocate the link.

Table 1 Fall 2015 ECU race/ethnicity data (undergraduate, graduate, medical, and dental programs all included) (<https://ipar.ecu.edu/wp-content/pv-uploads/sites/130/2020/01/Fact-Book-15-16.pdf>)

Race/ethnicity	Count	Percent
American Indian or Alaska Native	191	0.68
Asian	806	2.85
Black or African American	4497	15.90
Hispanic of Any Race	1599	5.65
Native Hawaiian/Pacific Islander	35	0.12
Non-Resident Alien	245	0.87
Race and ethnicity unknown	730	2.58
Two or more races	782	2.76
White	19,404	68.59
Total	28,289	

Traditionally, we have a larger fraction of White students in our introductory physics class than the university in general. This is a common pattern in physics across institutions

In this previous work, network properties under consideration were limited, and the changes seen over the course of the semester were described in terms of grades and group sizes. In this work, we describe these networks using a more robust set of network analysis tools.

Data

These data were collected in two introductory calculus-based physics courses taught in Fall 2015 and Spring 2016. Both courses were taught by the same instructor (the lead author of this paper: SFW) at East Carolina University (ECU). In both semesters, the instructor employed a group-work focused pedagogy in the daily class period. Students were expected to work in groups significantly more than 50% of the class time on problems or tutorials that were either of the instructor's design or were part of an established curriculum such as the Tutorials in Introductory Physics (McDermott and Shaffer 2002). Incidentally, one of the co-authors (TMS) was one of the students in these courses.¹ The physics courses are the standard Physics I (mechanics) and Physics II (Electricity, Magnetism, and Optics). ECU is a PhD granting institution in the southeastern United States with a strong regional recruitment presence.

Network data is based on student self-reports rather than a direct observation protocol. We argue that this is an authentic method for determining social connections, however, we do note that there are biases inherent in people which affect this data. Gender bias (McCullough et al. 2019) and racial bias (Cochran et al. 2019) are well documented in learning settings and are undoubtedly unconscious factors which bias individuals in our sample as they are reporting the others that they are working with. It is a source of systematic error that we do not attempt to quantify here. The institution's race/ethnicity profile is given in Table 1, and the gender breakdown² is given in Table 2 (<https://ipar.ecu.edu/wp-content/pv-uploads/sites/130/2020/01/Fact-Book-15-16.pdf>).

¹ Data was deidentified using a salted hashing algorithm before TMS worked with this data. We do not know which node he is in these networks.

² At the given time, the institution only collected male/female gender data.

Table 2 Fall 2015 ECU gender data broken down by program (<https://ipar.ecu.edu/wp-content/uploads/sites/130/2020/01/Fact-Book-15-16.pdf>)

	Total	Percent
Undergraduate		
Male	9599	41.66
Female	13,440	58.34
Graduate		
Male	1630	34.45
Female	3101	65.55
Medical		
Male	161	51.44
Female	152	48.56
Dental		
Male	103	50.00
Female	103	50.00
Total male	11,493	40.63
Total female	16,796	59.37

Traditionally, we have a smaller fraction of Female students in our introductory physics class than the university in general. This is a common pattern in physics across institutions

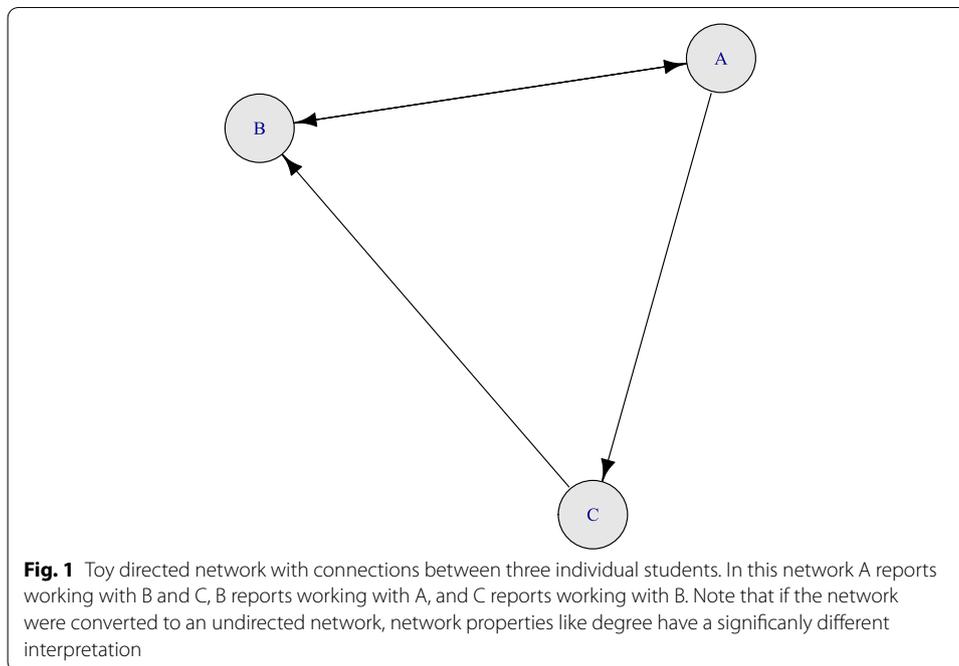
In the Fall semester, we had $N = 44$ students who took all of the assessments and $N = 36$ students in the spring semester. As these were consecutive courses, and there are multiple sections taught by different instructors at the institution, it is commonplace for students to change from one instructor to the other due to schedule conflicts or personal preference. At the time this data was collected, the instructor for the section being studied (SFW) was the only instructor in the ECU Physics department who used group exams. Since the time of this writing, other instructors have integrated group exams into their classes. There were $N = 22$ students who took both classes with the same instructor (SFW), and appear in the networks for both semesters.

There were 4 exams in each semester. Each had a multiple choice portion and a free response portion and had a 75 minute period to complete each portion of the exam. Students took the exam over the course of two class periods. During the first class period, students completed the exam on their own and turned in all exam materials. Then on the second day, students completed the exams in groups with fresh answer sheets (they were given the same problem sheets as the individual portion). For the group exam, the classroom environment was a large room with multiple round tables so that students could easily work together, and there were white-boards around the room so that students could work parts of the problems out. The fourth exam was also the course final exam. For the final exam, the individual exam and the group exam were consecutive, and the exam was written to be short enough so that it could be completed twice during the 150 min exam period.

Methods

Creating networks

An active, collaborative classroom is a natural setting for students' social connections to manifest themselves. We created two prompts to allow students to report who they



worked with and how closely. The first prompt was: “On this exam I *mostly* worked with...” and the second prompt was: “On this exam I *sometimes* worked with...” Students were free to choose what indicated these different levels of collaboration. If they asked, the researcher/instructor validated that they could and should apply their own definition as they saw fit. We had hoped to use the second prompt to generate multi-relational network data. However, we found that students tended to not use the second prompt. Indeed, subsequent interviews with students in these courses indicate that many did not see this addition of a second prompt was meaningful to them (Carr et al. 2018). In a few cases, students would simply mark every student on the second prompt only, which is too general to be informative about the nature of the collaborative relationship between two individuals. The number of edges that specific answers to this prompt would add was so small that we chose to ignore the *sometimes* prompt entirely in this analysis.

Our method for parameterizing social networks from self-reported data is drawn from the method developed by Brewe et al. (2012). We utilized a directed network framework, rather than an undirected network as it would lose information about relational reciprocity. For example, if we have three students, A, B, and C, and the following relational data, A reported working with B and C, B reported working with A, and C reported working with B, we would get the network shown in Fig. 1.

In the subsequent sections we discuss the methods that we use to describe these networks in the context of a classroom. The measures that we will use break down into several categories:

1. Global network statistics - these are all single number measures, such as the number of nodes in the network.
2. Node property measures and distributions - for example, we will look at the degree centrality distribution.

3. Network partitioning methods including community detection and blockmodeling.

In the networks that we consider, we will remove vertices representing students that don't take every exam. A small fraction of students (historically, < 5% for the instructor of record) drop these courses or stop attending class at some point after taking the first exam.

Global statistics

We will characterize the exam networks for each semester by several measures that are single-number summaries of the network as a whole. In addition to the number of vertices and edges, we consider density, reciprocity, transitivity measures, the average network distance, and degree assortativity. *Density* is the fraction of total to possible edges (Prell 2012). It is often reported in PER network studies, but is less useful for comparing networks of different sizes. *Reciprocity* is the fraction of named links that were returned (Prell 2012), which has been found to change over the semester as grade disparities emerge (Wolf et al. 2017).

Clustering is a more size-stable measure of connectedness than density, and is considered through two coefficients. *Transitivity* compares the number of triangles to the number of connected triples, and is the average probability that two students will be linked if they both link to a third student (Newman 2003) (“the friend of my friend is my friend”). The second is the local definition of clustering coefficient by Watts and Strogatz (1998), where each node's clustering coefficient is the fraction of its neighbors' possible edges that actually exist. This clustering coefficient is then averaged over all edges to give a global network score. This average local clustering coefficient will be higher if students tend to form tightly connected “pods,” versus seeking a more diffuse set of partners who may not talk to each other.

The *average vertex-vertex distance* (Newman 2003) gives a measure of how strong the “small world” effect is for the networks. On the time-limited task of exams, this distance may indicate how easily information about how to work the problems circulates through the network. For unconnected or weakly connected networks of N nodes, it can be computed either by only counting existing paths, or by counting “missing” paths as having length $N + 1$. Comparing both values gives a sense of how much network distances are skewed by the lack of paths between components. Finally, *assortativity of degree* (Newman 2003) is the correlation coefficient between node degree (retaining edge direction), and is typically positive for social networks, showing a tendency for well-connected students to preferentially talk to each other.

It can be inappropriate to compare some of these statistics, such as density, for networks of different size. However, we will restrict our comparisons to networks for a single semester, and have cleaned the raw network data to include people who took all of the group exams. Therefore, all networks being compared within a semester are the same size, and these statistics can be compared directly.

Node property measures and distributions

One way to investigate network change is by looking at how the centrality distributions relate for each exam network. We focus on four centrality measures that are the most common in educational network studies (Saqr and López-Pernas 2022):

Degree The number of edges coming into (in-degree) or leaving (out-degree)

- each node (Freeman 1978). High in-degree is one measure of popularity.
- Betweenness** The total number of (directed) shortest paths that pass through a particular node. High betweenness has been described as advantageous due to an “information broker” position (Prell 2012), but can also indicate an unfavorable state of being marginal to multiple groups (Dawson 2008).
- Closeness** The reciprocal of the sum of the length of the shortest paths between the node and all other nodes in the graph (Freeman 1978), a measure of the extent to which a student is “in the thick of things” versus on the edge of collaborations.
- Eigenvector** The component of the eigenvector related to the largest eigenvalue for the adjacency matrix of the directed network (Bonacich 1987). Eigenvector centrality encompasses the idea that not just the number but the relative popularity of your collaborators can increase your centrality score.

We used the programming language R (R Core Team 2020) and the `igraph` package (Csardi and Nepusz 2006) to compute these centrality measures for our networks. While it is not generally expected for these measures to strongly correlate for a single network, we are interested in understanding how node centrality changes as the network develops. We will use the directed versions of betweenness and eigenvector centrality calculations, as they reflect the available information about reciprocity of ties, which has been found to segregate by grade over the semester (Wolf et al. 2017). We will also choose inward directed measures (in-degree and in-closeness) as each of these scores for a person do not depend on the relationships reported by that person. For example, my out-degree is simply how many people that I reported working with, while in-degree is the number of people who reported working with me.

Network partitioning measures

The network literature is full of methods for partitioning networks (e.g., Newman and Girvan 2004). In general, these methods attempt to break vertices in a network into groups based on different criteria. One criterion is to detect communities: groups of nodes that are significantly more connected to other nodes in the same community than those outside. We will use the Girvan-Newman or `edge-betweenness` algorithm for this purpose (Newman and Girvan 2004). Because the name `edge-betweenness` gives the reader insight as to the network properties whereby this algorithm makes communities, we will use that name for the rest of this paper. However, community detection algorithms don't give us much insight into the social roles that individuals are playing within a network. For this, we will look at structural equivalence. Structural equivalence partitioning methods focus on how vertices connect to other vertices, and groups vertices if they share similar linking behavior. For example, the outer nodes of a star-shaped network would all be structurally equivalent to each other and thus form a block, even though none of them directly connect to each other. Structural equivalence algorithms are good at doing a positional or role analysis in social networks. We will use the `CONCOR` algorithm (Breiger et al. 1975) for this purpose.

Edge betweenness

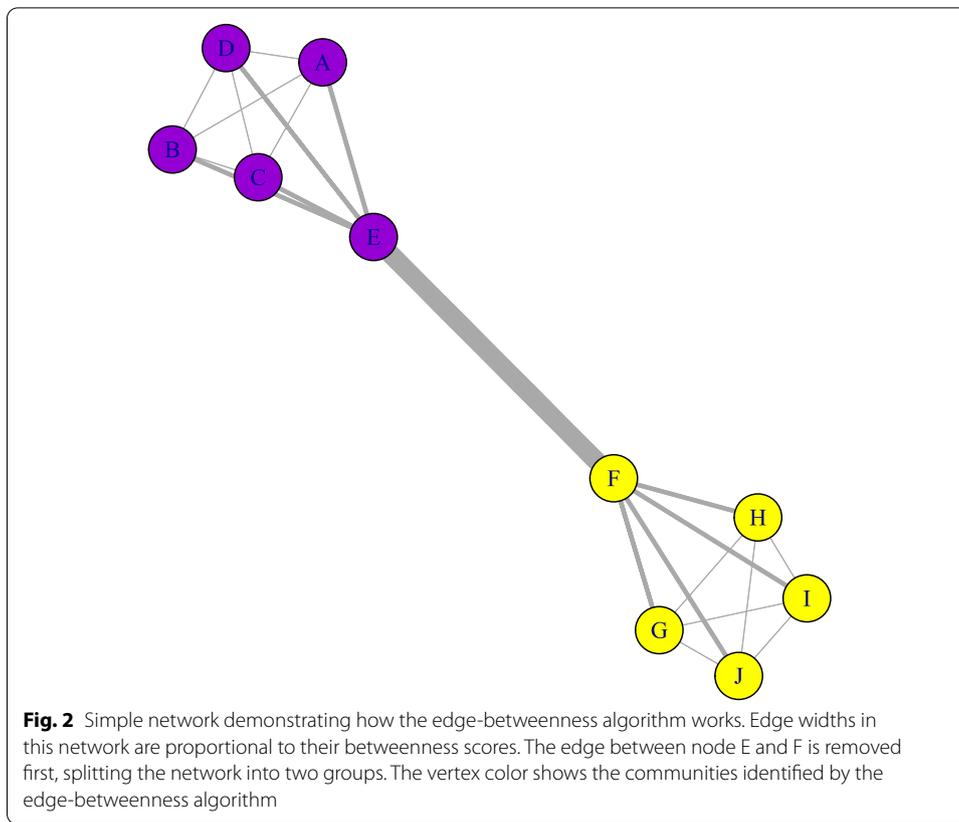
Edge betweenness determines community structure by using a divisive process, rather than an agglomerative process. We have chosen this algorithm for several reasons that are more fully discussed previously (Wolf et al. 2016) and summarized below. First, the designers promoted this algorithm for smaller networks as they found that it was not computationally feasible for large networks on the order of 10^5 nodes.³ As our networks are not nearly this size, we don't have to worry about computational limits. Furthermore, when this algorithm was compared to other community detection algorithms such as the walk-trap algorithm (Pons and Latapy 2005), we found that the communities determined were either identical or created fewer communities within the network. We also found that the number of communities detected by edge-betweenness tended to match the number of tables that students worked at in the classroom where they took the exam (Wolf et al. 2016), suggesting a good match between the edge betweenness criterion and observed classroom structure. We utilized the `cluster_edge_betweenness` function in the `igraph` package (Csardi and Nepusz 2006) of the R programming language (R Core Team 2020). The only required input to this function is the network object, however it can be configured to treat a directed network as an undirected network or change the weights of edges. We used this algorithm in the default configuration, allowing it to account for the information embedded in the directional network ties.

The edge betweenness algorithm iteratively removes edges in order to maximize the network property of modularity (Newman and Girvan 2004). The modularity is a measure of how likely an edge connects two members of the same community. So, if student A and student B are connected, the modularity is the probability that student A and student B are in the same community. The edge betweenness algorithm works according to the following process:

1. Calculate betweenness scores for all edges in the network and the network modularity.
2. Remove the edge that has the highest betweenness score (in the event of a tie for the maximum score, choose one of them at random and remove it).
3. Recalculate the betweenness scores and network modularity.
4. Repeat steps 2 and 3 until you can determine the global maximum of the network modularity. (Worst case, this will be repeated until no edges remain in the network.)

By removing edges with the highest betweenness, the algorithm removes the edges which tend to move between communities as the toy network in Fig. 2 demonstrates. Once the maximum modularity is determined, vertices are grouped into communities by grouping nodes that are connected with each other by any path. For example, in the toy network considered previously, if the edge between vertex I and J were removed, the community structure would remain the same. It should be noted that while the toy network is an undirected network, the edge betweenness algorithm can be applied

³ At least, given the computational limitations of 2004. This limit might be raised given modern computers implementing the algorithm with parallel processing.

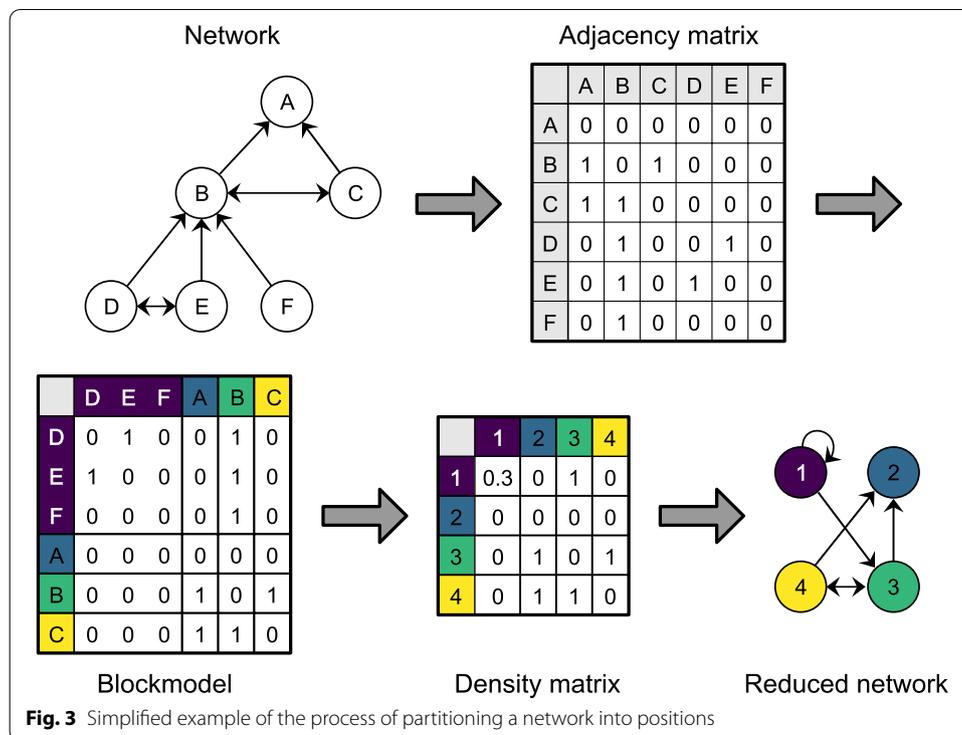


to directed networks as well, and betweenness scores are well-defined for directed networks.

CONCOR

CONCOR (CONvergence of iterated CORrelations) is an algorithm for positional analysis developed by Breiger et al. (1975). We selected CONCOR because of its prominence (Wasserman and Faust 1994) and continued use (Luo et al. 2014) as a structural equivalence tool, and because more generalized approaches and regular equivalence methods encounter more empirical trouble finding a global solution (Ferligoj et al. 2011). CONCOR splits the network into two groups by correlating columns in the adjacency matrix (after isolated nodes have been removed). It then correlates the columns in the correlation matrix and repeats this process until all values are ± 1 or the maximum number of iterations are reached.⁴ Then it groups the nodes into the +1 block and the -1 block. The algorithm can then be repeated on each block as often as the user wishes leading to 2^n partitions, where n is the number of times the algorithm is repeated. We have implemented the CONCOR algorithm in R using the `concorR` package (Suda et al. 2020).

⁴ If the max number of iterations are reached, no split is found and the network is not partitioned.



Network visualization

A key feature of CONCOR, or any other network partitioning algorithm, is the abstraction of the network into inter- and intralinked positions. After the adjacency matrix has been permuted by CONCOR, each block in the matrix—on or off the diagonal—can be thresholded to 0 or 1. The most common threshold is the whole-network edge density—or normalized degree. The density can be compared to the density of each block (total number of links in the block divided by nm for an $n \times m$ off-diagonal block, or divided by $n(n - 1)$ for an $n \times n$ diagonal block). With each block set to 0 or 1, the network is simplified to a reduced *adjacency matrix*, which can be plotted as a network. Figure 3 shows an example of these steps.

Longitudinal analysis

When longitudinal data is available for a network, CONCOR can be run on each “snapshot” of the network independently, or it can use the entirety of the data to calculate a single set of positions. In the latter case, it stacks all r available adjacency matrices into a single $rN \times N$ matrix before correlating the columns. Both sets of results—CONCOR partitions generated from single networks or a time-connected set—are included below. In either case, the nodes assigned to a position can change greatly between time points, and this is often not obvious from the sociograms or reduced network plots. To evaluate the stability of students’ CONCOR partition, we also include alluvial diagrams (Rosvall and Bergstrom 2010). These show the “flow” of membership in a category (here, CONCOR partition) at successive times for the same entities.

Results and discussion

Fall 2015

Global and node-level measures

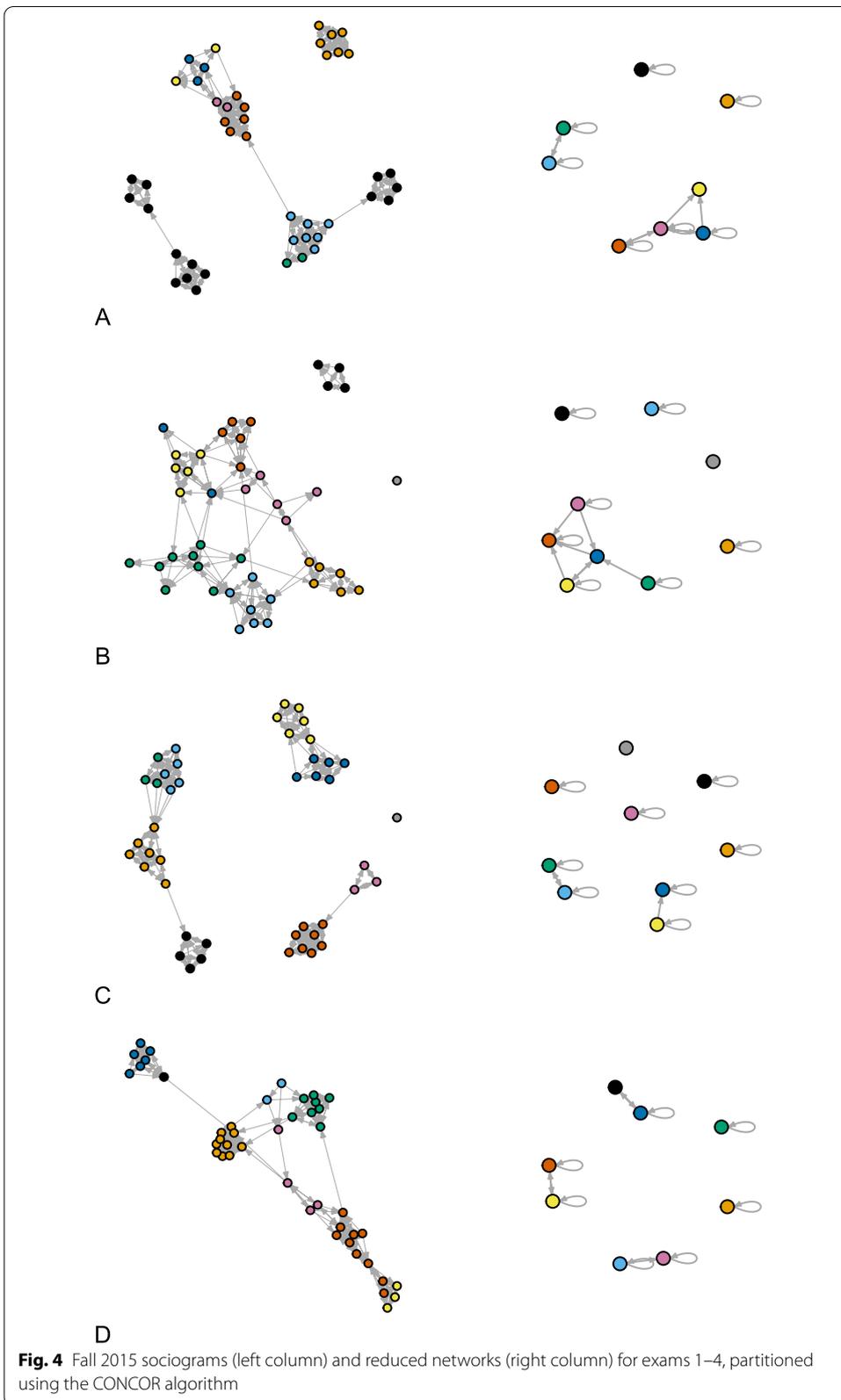
In the Fall 2015 semester, 48 students were enrolled in the course (Physics I) on census day. Three students withdrew, and one student, who did not take all of the exams, received a failing grade. As a result, our networks for Fall 2015 have $N = 44$ nodes. Table 3 shows summary statistics for the fall semester. The left hand column of Fig. 4 shows the sociograms for each of the exam networks, with nodes colored by their CONCOR block membership. From the first test to the second, there was a sizable drop in the number of edges and reciprocity of named links. This corresponded to a lower density and average degree. The second exam also had a notably lower transitivity, though its average local clustering coefficient (AvgCC) remained comparable to the others. This contrast may occur because the local clustering coefficient tends to heavily weight low-degree nodes (Newman 2003), of which there were more on exam 2. Exams 2 and 4 had the highest average vertex-vertex distance ignoring disconnected node pairs (AvgDist) but the lowest vertex-vertex distance when disconnected node pairs are included (AvgDistUC). This occurs because exams 2 and 4 are (at least weakly) connected networks, while exams 1 and 3 have several unconnected sub-components. Finally, the degree assortativity varies widely across exams, being high for the first and last exams, a moderate value for the third, and effectively zero for the second test. Broadly, exam 2 seems to have scattered nascent social structure, which re-established itself later in the semester.

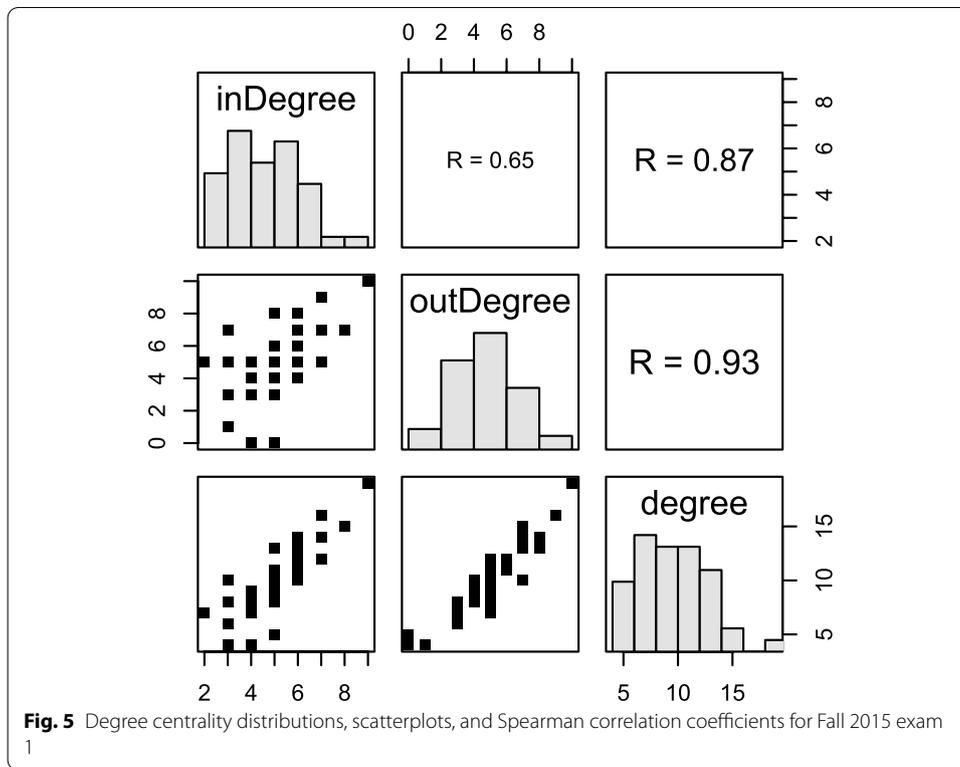
In addition to comparing the values of the network measures for each of the exams, we also analyzed centrality distributions for each exam network and explored how they evolved over the semester. In general, undirected versions of each statistic correlated well with their directed versions for the same exam. As an example, we present the different types of degree distributions for Fall 2015 Exam 1 in Fig. 5 as well as how they correlated with each other. These distributions were frequently not normal, so we used the Spearman correlation in this paper. We should note that the out-degree distribution is peaked in the middle, which is not common for network degree distributions. This is likely due to the fact that the tables that the students worked at had eight seats. The fact that this distribution is bell shaped is going to be strongly influenced by the fact that the average number of students sitting at each table was about six (44 students sitting at seven tables) and students were observed generally interacting with everyone at their table. Within a single network, it is not surprising that related centrality measures correlated with each other, and the correlation observed for the degree family of centrality statistics continued for the other families of centrality measures.

Table 3 Summary statistics for the fall semester ($N = 44$ nodes)

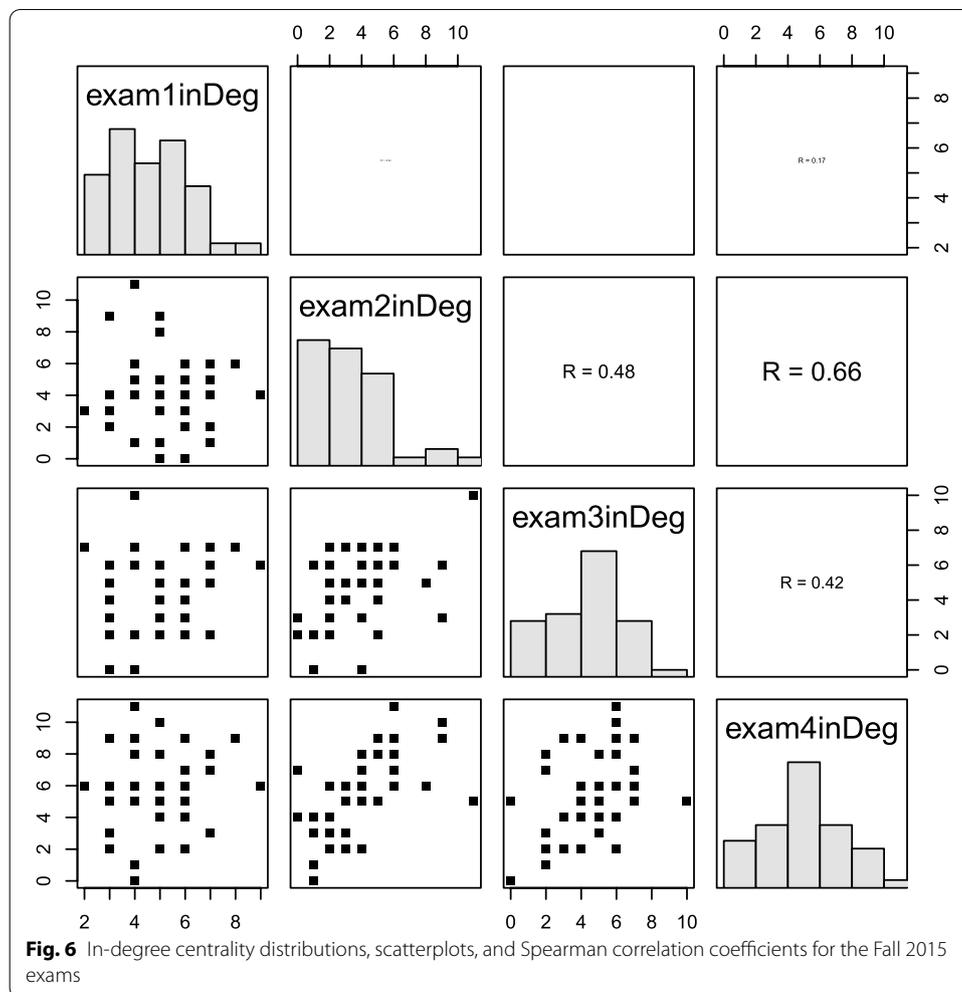
Nedge	Reciprocity	Density	AvgDeg	Transitivity	AvgCC	AvgDist	AvgDistUC	Assort
223	0.83	0.12	10.14	0.83	0.30	2.03	33.35	0.46
172	0.63	0.09	7.82	0.54	0.35	4.34	16.65	0.03
211	0.78	0.11	9.59	0.81	0.30	2.04	33.89	0.32
247	0.75	0.13	11.23	0.79	0.33	4.25	11.33	0.53
399	0.84	0.21	18.14	0.61	0.22	2.31	2.31	0.19

Each row shows the network statistics for tests 1 through 4 and the last row shows the statistics for the aggregate network





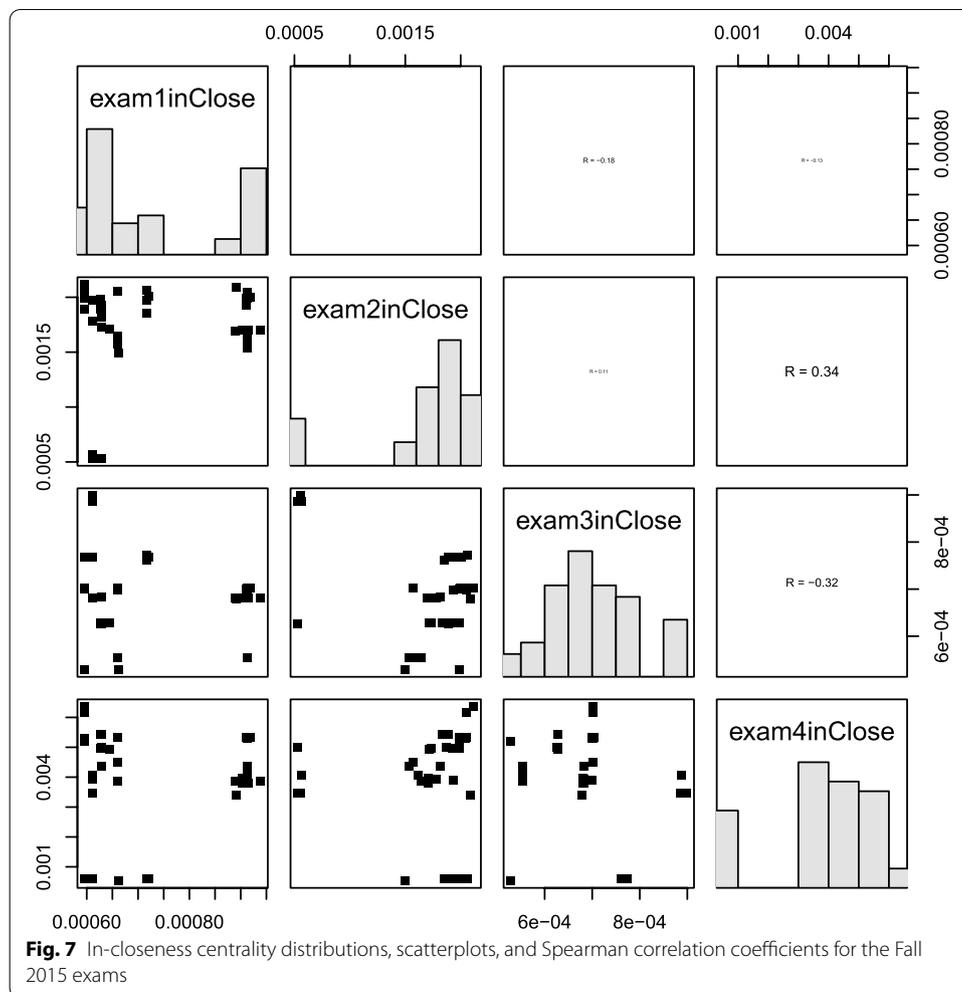
We are also interested in how centrality lasts for nodes in evolving networks. In particular, are highly central nodes in early networks also highly central nodes in later networks? As we discussed earlier, we will focus on directed or inward measures of network centrality. In Fig. 6, we show the distributions, scatterplots, and correlations for in-degree centrality for each of the exams in Fall 2015. The correlation between exam 1 in-degree and any other exam in-degree was small ($R = 0.17$ was the largest). But for subsequent exams, the correlation became stronger. In Fig. 7, we show the distributions, scatterplots, and correlations for in-closeness centrality for each of the exams in Fall 2015. None of the correlations were significant ($R = 0.34$ was the largest correlation observed) and some correlations were negative. A significant fraction of nodes had a notably small in-closeness relative to the group, making the distributions bi-modal the correlation coefficients more difficult to interpret. In Fig. 8, we show the distributions, scatterplots, and correlations for directed eigenvector centrality for each of the exams in Fall 2015. The correlations for these distributions were similar to the in-closeness distributions in that they were driven by bi-modal distributions in the centrality scores. There was a notable correlation ($R = 0.48$ between exams 2 and 4), but this was highly influenced by the large fraction of nodes with an eigenvector centrality of nearly zero. In Fig. 9, we show the distributions, scatterplots, and correlations for directed betweenness centrality for each of the exams in Fall 2015. The betweenness statistic returns (somewhat) to the pattern that we observed with the degree statistic. However, it is interesting to note that the maximum betweenness score varied by approximately a factor of 5–6 on the different exams (approximately 100 on exams 1 and 3 and 500–600 on exams 2 and 4). In all distributions, the mode betweenness score was zero, suggesting that the



correlation is due to the censored nature of the distributions. Another way of putting this is that in these classroom networks, most nodes were not very “between” regardless of the exam. However, there are also not a consistent set of students that are highly between that are driving the classroom collaboration networks during the fall semester.

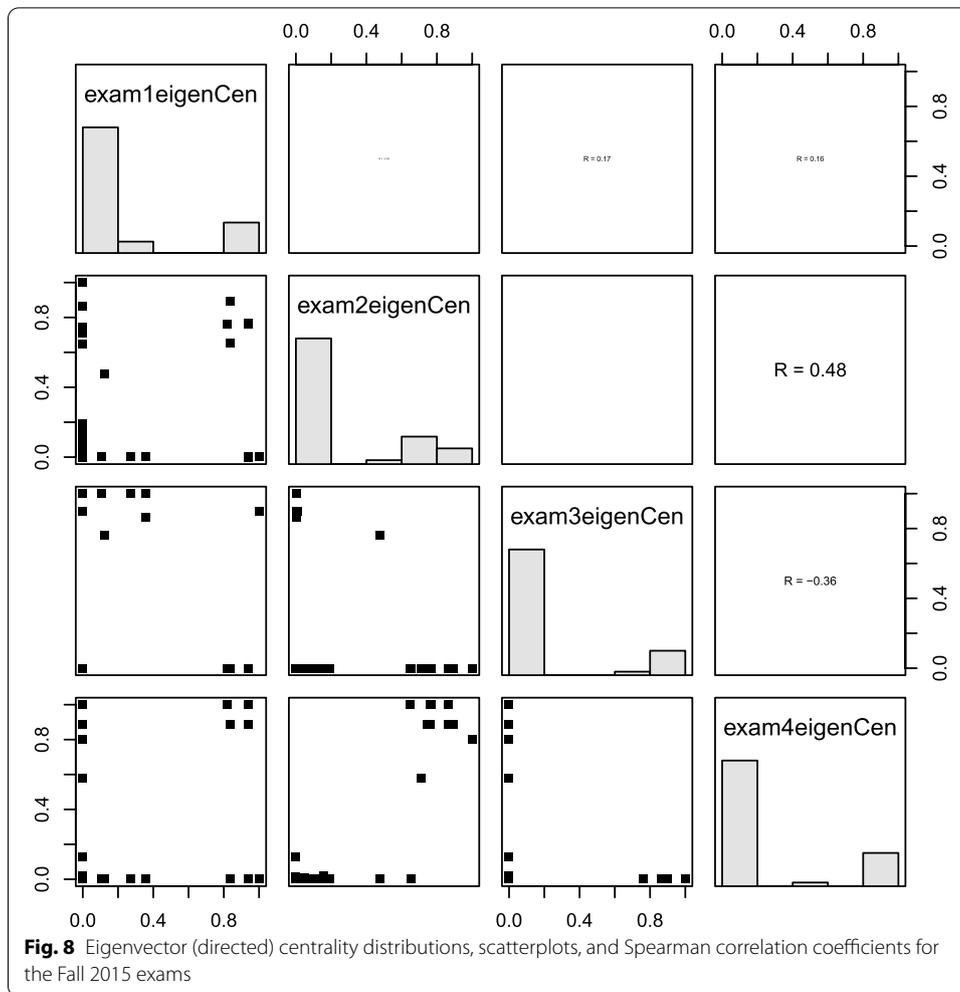
Network partitioning

We are also interested in looking at how network roles change over the course of the semester using CONCOR. Figure 10 illustrates the difference that can emerge in going from two to three CONCOR splits for the second exam in Fall 2015. The first and second positions split along fairly obvious lines: two subgroups which were not connected to each other in the first case, and two internally-dense subgroups with a smaller number of bridging links. The third position splits into a core group of five nodes and a two-node position of students who have no connections to each other, but are both peripheral to the core position. Finally, the fourth position splits into a dense group and a secondary group with only sparse links, either to each other or to the main group. We have



investigated CONCOR block membership with 3 splits for each of the exams in the Fall 2015 semester.

Block membership on exams 1 and 2 have elements that are common to community detection algorithms, for example, isolated groups form several blocks. But they also exhibit notable differences. For example, on exam 1 (panel A in Fig. 4) CONCOR splits the top bundle of nodes into 4 different blocks (node color is generated from each network’s CONCOR block, and does not persist from network to network). These two cases also show a behavior that is unlikely or impossible in most community detection methods: grouping together nodes which are loosely or entirely unconnected to each other, but which belong together because of their linking behavior with respect to another network position. 1 Fig. 11 shows the same networks with node colors based on CONCOR (left hand column) and edge-betweenness (right hand column) for each of the exams. In each case, the CONCOR splits can show marked differences in nodes compared to edge-betweenness. For example in Exam 4 (Fig. 11, row D), the green group identified by edge-betweenness is almost reproduced by CONCOR with one notable exception. There is a single orange node, connecting that group to the rest of the network. That student is performing a function different from the rest of the “green” group. CONCOR can



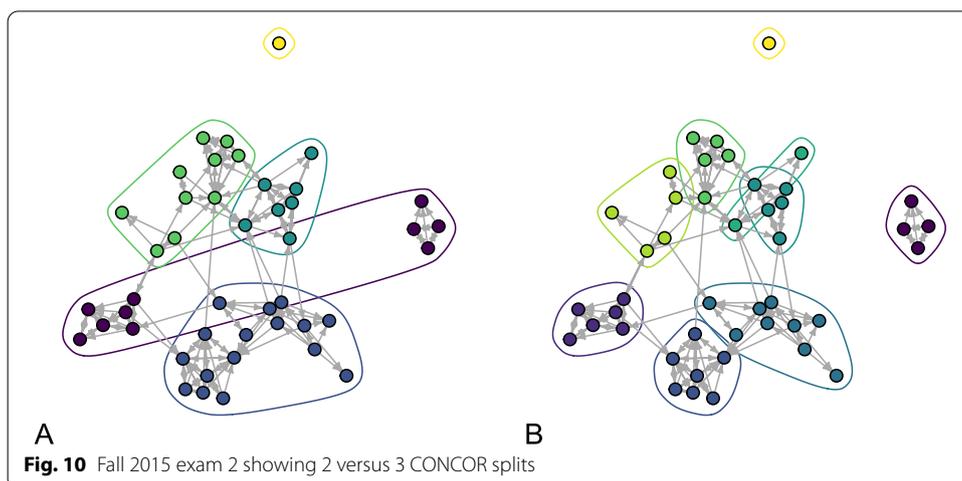
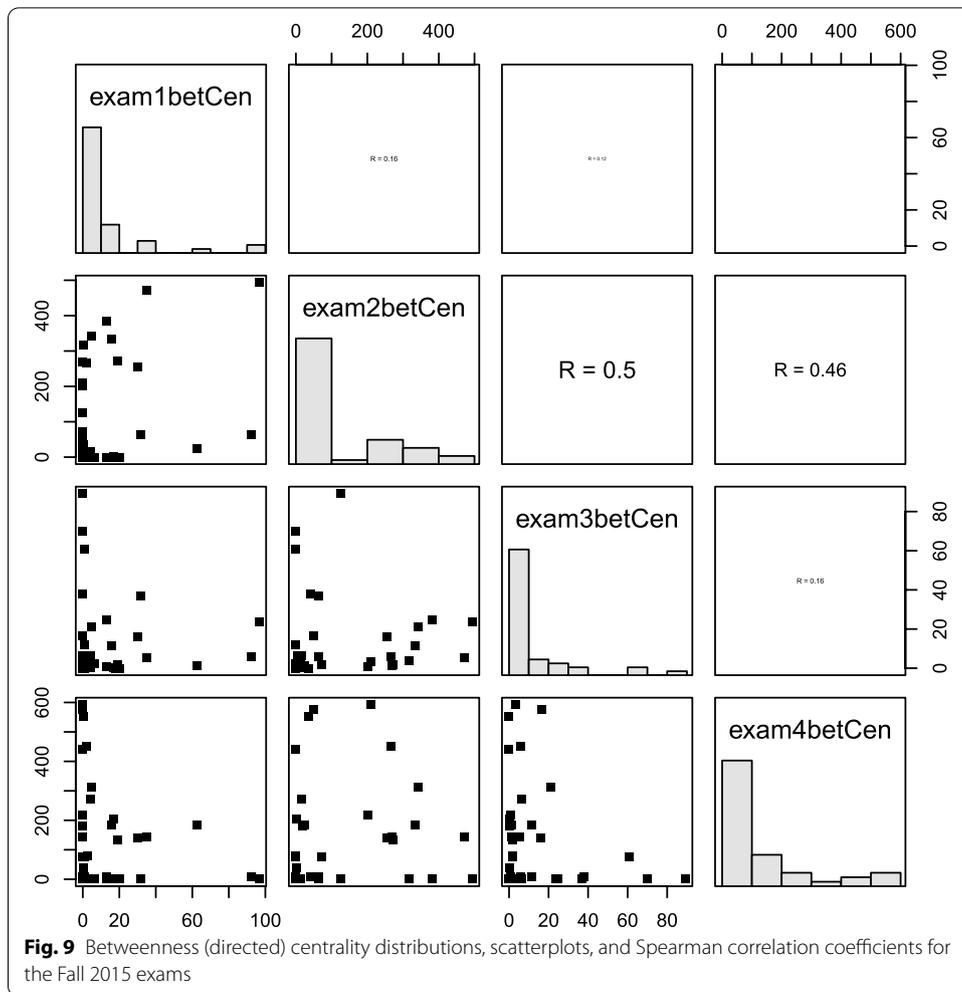
and often does detect clusters that are internally dense, but it can also highlight nodes that are visually part of a larger cluster but in fact are only peripherally tied to it. In the right hand column of Fig. 4, we present the reduced networks for these exams. We find that blocks are more connected on exams 1 and 2 than they are in exams 3 and 4 as evidenced by the number of inter-block connections.

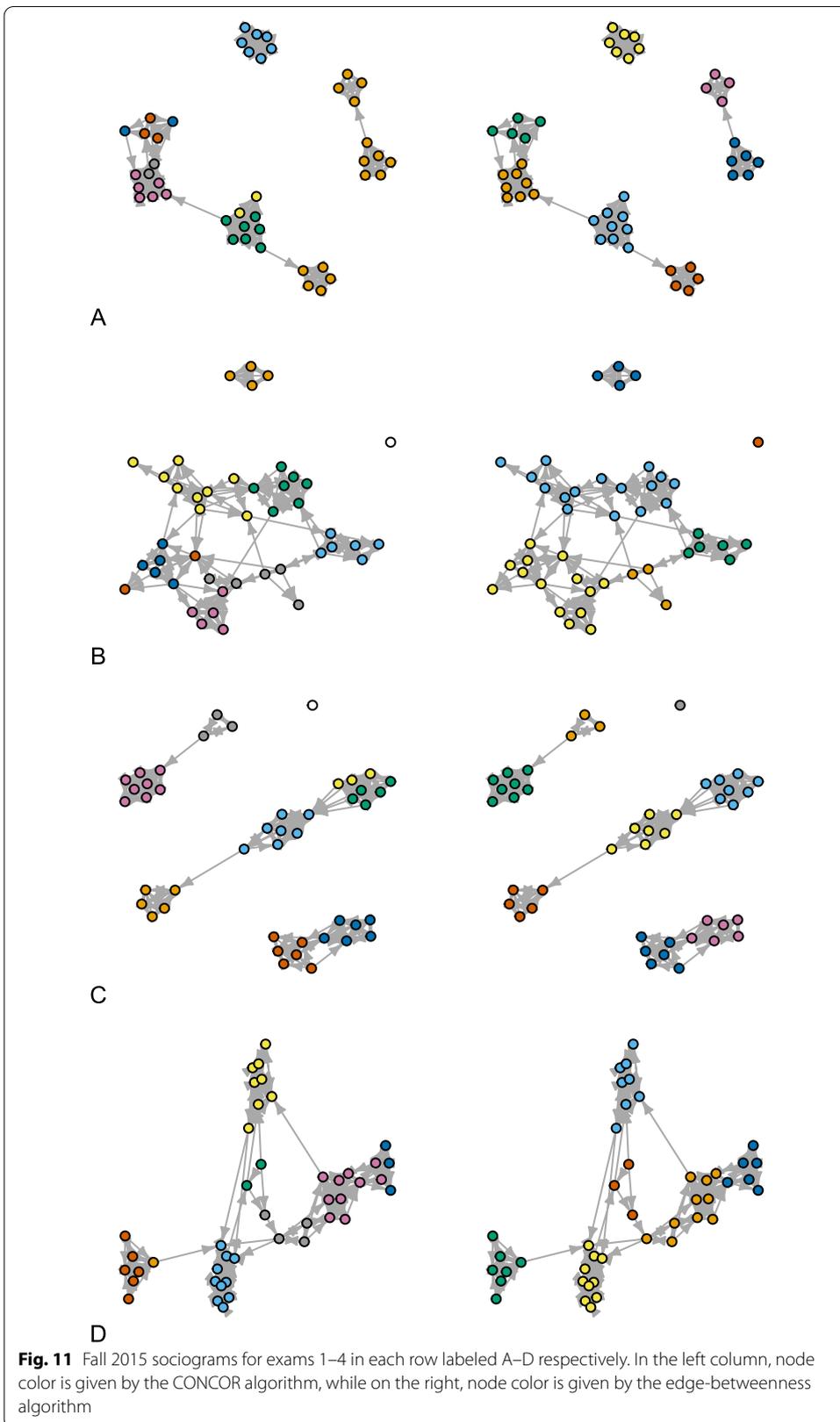
Finally, it is clear that the blocks found by CONCOR are not stable across exams during the fall as shown in the alluvial diagram (Fig. 12). This leads us to note a few things. First, the block number assigned by the algorithm is not significant—they just have to do with what block is the “easiest” to detach from the network. In general, nodes that are together in one block during an exam are not necessarily blocked together in subsequent exams, although a few cohorts of students stay together throughout the semester (for example, the band that goes from block 7 to block 5 to block 5 to block 1).

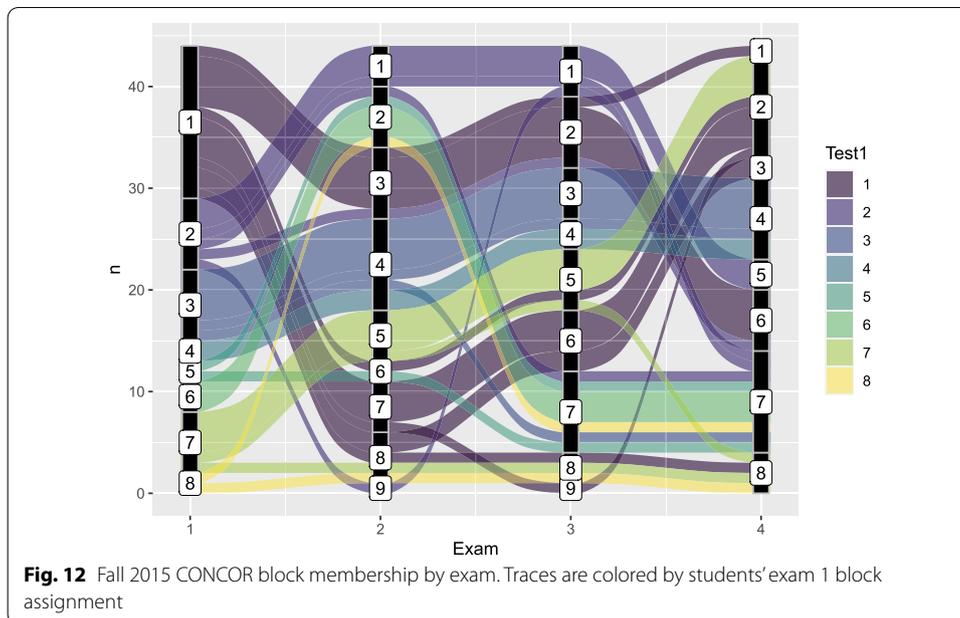
Spring 2016

Global and node-level measures

In the Spring 2016 semester, 36 students were enrolled in the course (Physics II) on census day. All of the students took all of the exams. Therefore, the networks for







Spring 2016 have $N = 36$ nodes. As stated previously, some of the students ($N = 22$) took the previous N course in the Fall 2015 semester. The left hand column of Fig. 13 shows the sociograms for each of the exam networks, with nodes colored by their CONCOR block membership. Table 4 shows summary statistics for the spring semester. By and large, the summary statistics were much more stable across exams than during the fall semester. One possible mechanism to explain this stability is that group exams were an unfamiliar event for all students in the Fall 2015 semester, but not so for the 22 students in the Spring 2016 semester who were in the Fall 2015 course. This added familiarity with group exams could have led to a more swift adoption of group exam collaboration norms. The number of edges was consistent over the first three exams, and then increased slightly on the fourth exam. As a result, the density was also stable for all four exams. The average degree was stable for the first three exams and then increased by approximately 1 for the fourth exam. The reciprocity increased from exam 1 to exam 2 by 9%, but other shifts between exams were smaller. There aren't notable differences between the fall networks and the spring networks based on these measures, and the global and local clustering coefficients were similar as well. Finally, the degree assortativity has the most variation across exams, being low for the first exam, spiking in the second exam, a moderate value for the third, and increasing again for the fourth test.

The centrality distributions for each exam network in the Spring 2016 semester exhibited some similar patterns to those found in the fall semester. As an example, we present the different types of degree distributions for Spring 2016 Exam 1 in Fig. 14 as well their Spearman correlations.

What is more interesting about this analysis is investigating how centrality “lasted” in the Spring 2016 semester. Similarly to the fall semester, and somewhat surprisingly given the fact that slightly more than half of the class was familiar with the group

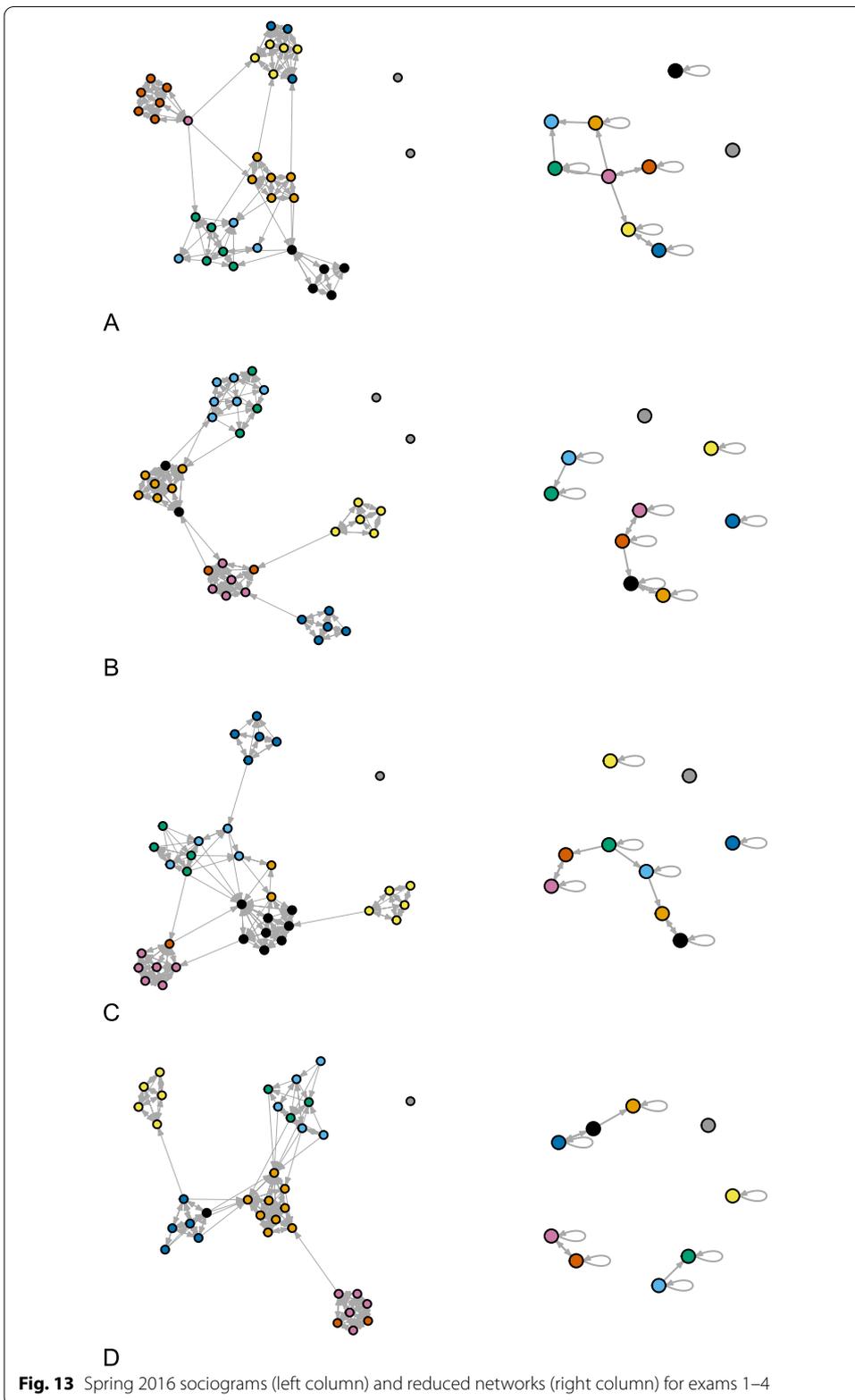


Table 4 Summary statistics for the spring semester ($N = 36$ nodes)

Nedge	Reciprocity	Density	AvgDeg	Transitivity	AvgCC	AvgDist	AvgDistUC	Assort
160	0.65	0.13	8.89	0.73	0.36	2.79	18.16	0.04
162	0.74	0.13	9.00	0.78	0.31	3.60	15.20	0.44
161	0.71	0.13	8.94	0.68	0.34	2.74	17.79	0.10
180	0.77	0.14	10.00	0.71	0.32	1.88	24.46	0.20
258	0.78	0.20	14.33	0.62	0.24	2.39	4.25	0.21

Each row shows the network statistics for tests 1 through 4 and the last row shows the statistics for the aggregate network

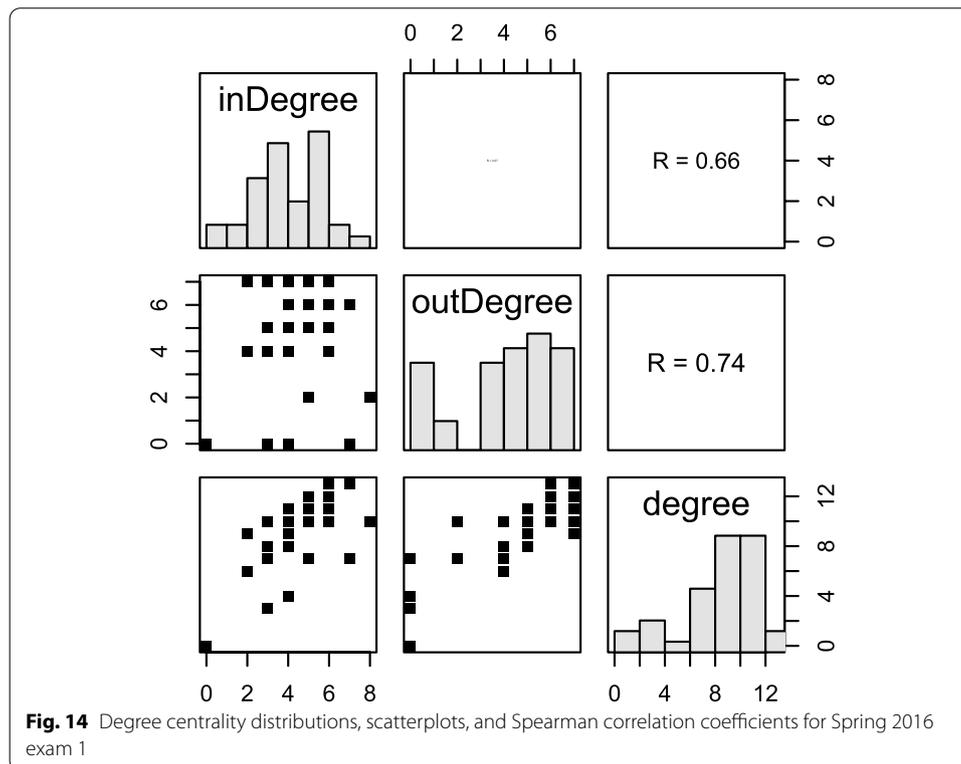
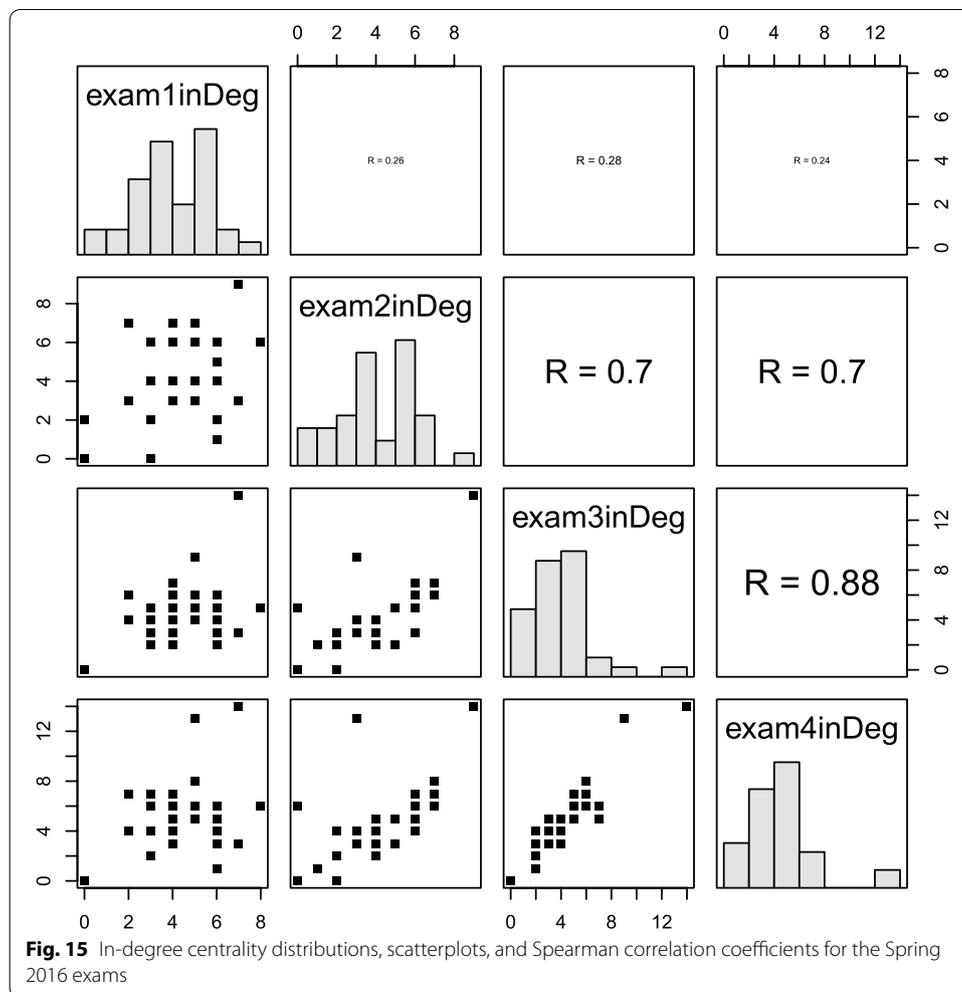


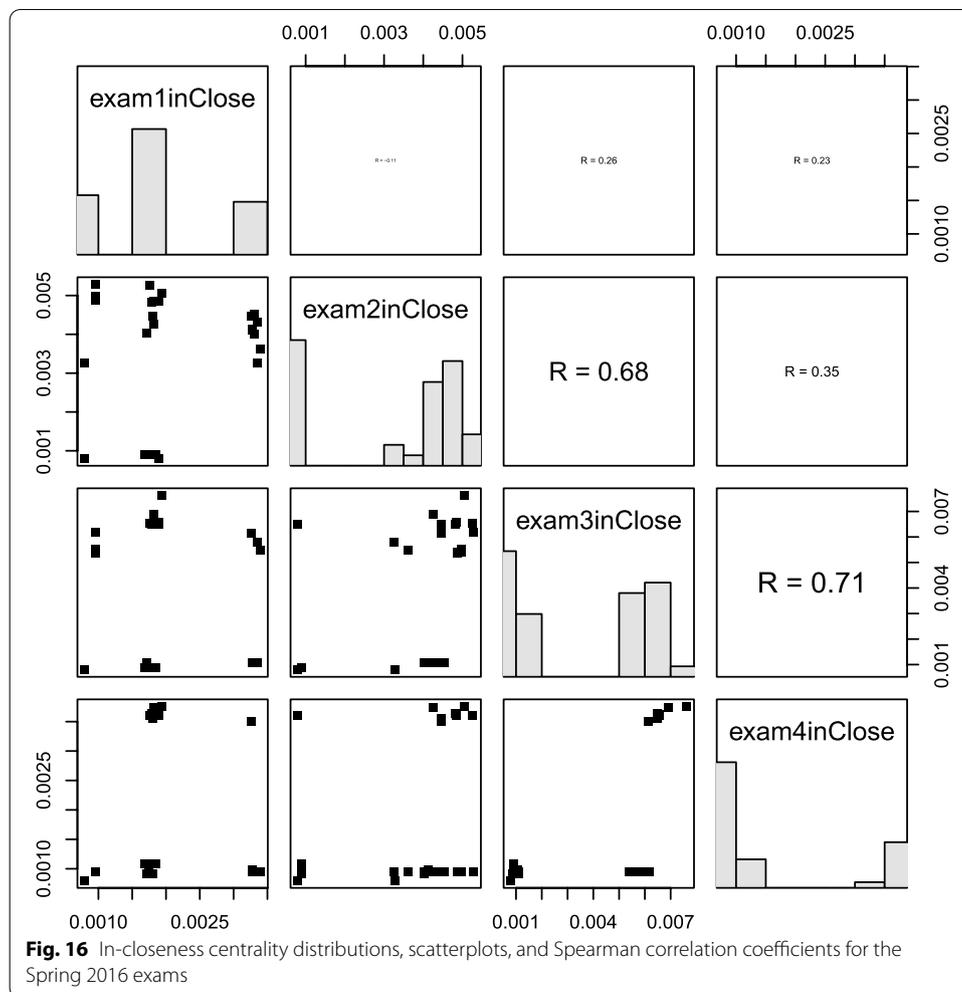
Fig. 14 Degree centrality distributions, scatterplots, and Spearman correlation coefficients for Spring 2016 exam 1

exam paradigm, we observed that the centrality scores in first exam network did not correlate with centrality scores on future exam networks. In Fig. 15, we show the distributions, scatterplots, and correlations for in-degree centrality for each of the exams in Spring 2016. Here, the trend observed in the fall is amplified—correlations between exam 1 and other exams were small ($R = 0.28$ between exams 1 and 3 was the largest correlation score), and were stronger between exams 2-4, ranging from $R = 0.70$ to $R = 0.88$. In Fig. 16, we show the distributions, scatterplots, and correlations for in-closeness centrality for each of the exams in Spring 2016. For closeness, we observed a similar pattern to the degree statistic: Exam 1 did not correlate strongly with other exams, and exams 2-3 correlated more strongly with the subsequent exam ($R = 0.68$ for exams 2 and 3 and $R = 0.71$ for exams 3 and 4). We also noticed that the closeness statistic was bi-modal for exams 2-4. Figure 17 shows these plots for directed eigenvector centrality. Again, exam 1 does not correlate with other exams, and exams



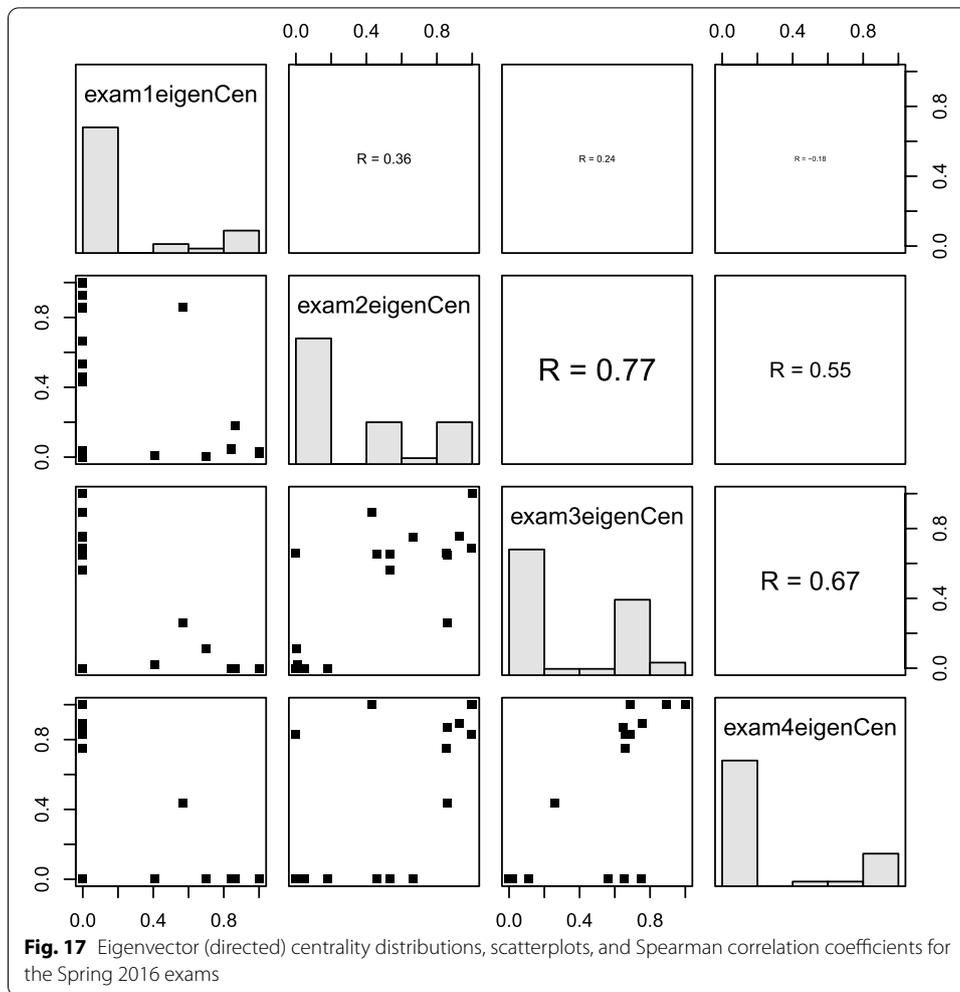
2–4 correlate with each other, especially the subsequent exam ($R = 0.77$ for exams 2 and 3 and $R = 0.67$ for exams 3 and 4, while $R = 0.55$ for exam 2 and 4). These distributions are still bi-modal, but are not as extreme as the closeness distributions. Figure 18 shows directed betweenness centrality for each of the exams in Spring 2016. The betweenness centrality does not follow the pattern established for the other centrality statistics. In general, all of the correlations were weak, with the exception, of Exam 3 and Exam 4. During this semester it is important to note that a small number of students (one of whom was TMS) were highly active in engaging their classmates in the last two exams. Even after many other students (including those in the group they worked with most on other days) had decided their work was complete, turned in their exams, and left, this group of student continued to engage with the rest of the class asking questions, getting ideas, and sharing their own answers to the problems. It is reasonable to assume that many students identified at least one from this group due to this gregarious behavior.

The pattern that we have described for the centrality distributions is echoed in our analysis of CONCOR block membership. We noticed that there was a re-numbering of the blocks between exam 1 and exam 2 on the alluvial diagram presented in Fig. 19, but



the groupings of nodes in blocks was relatively stable. After the first exam, this numbering of blocks was more stable than in the fall semester.

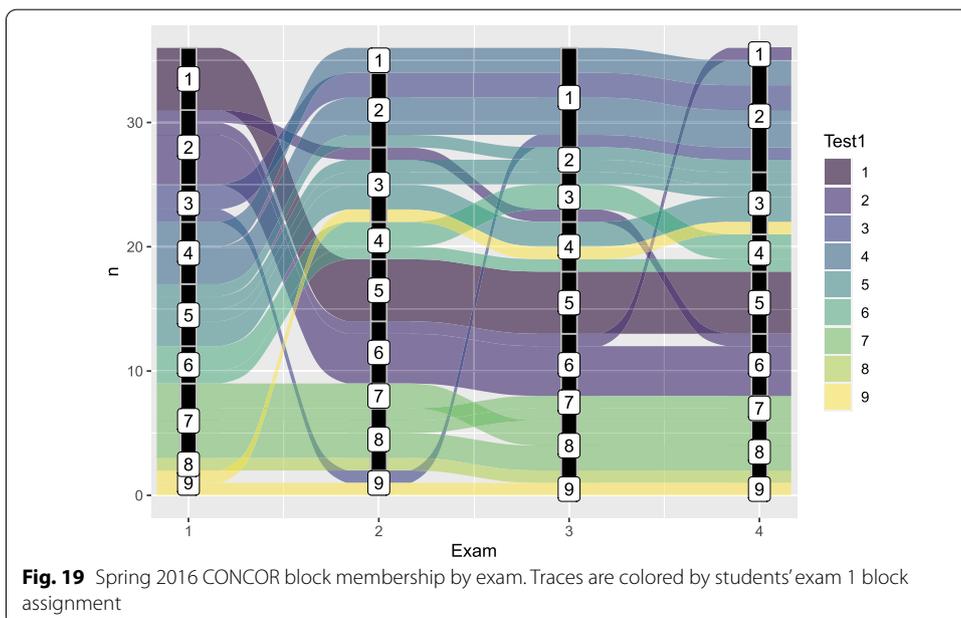
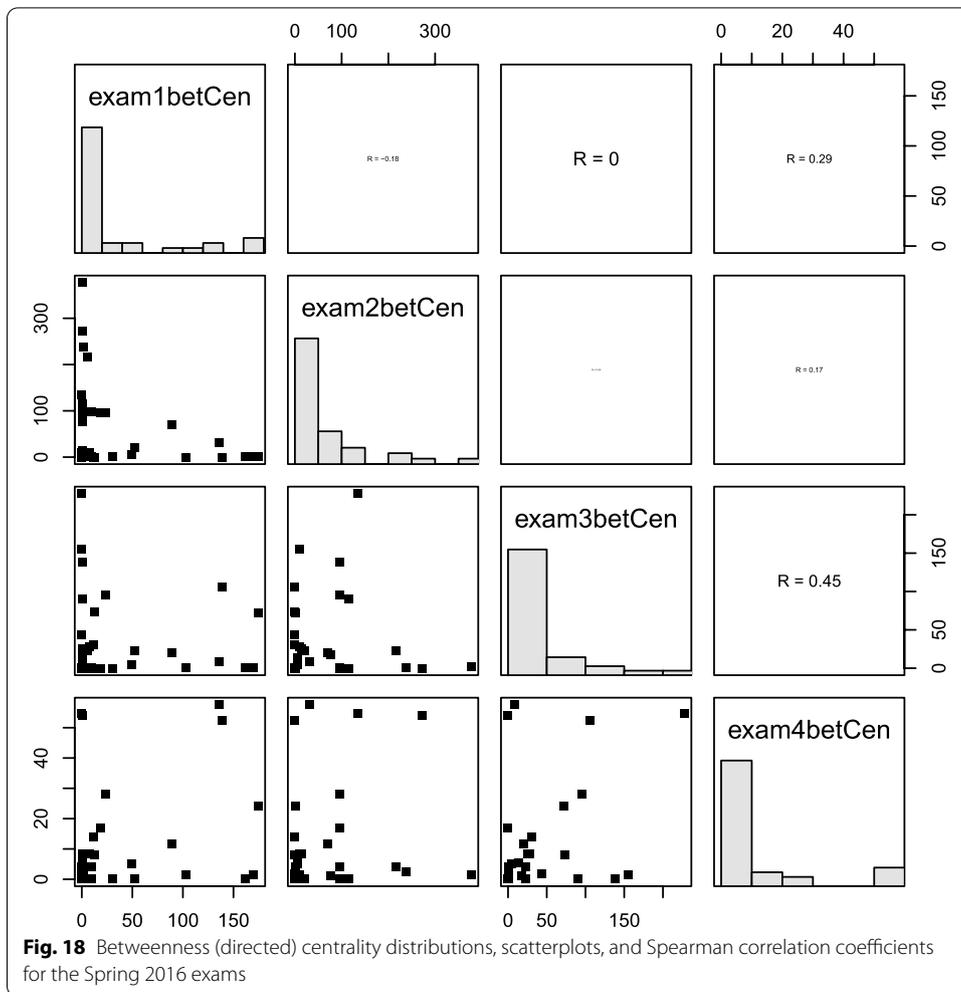
We also notice a more striking difference between the CONCOR blocks and the edge-betweenness communities in the spring networks. Sociograms for each of the exams are shown in each row Fig. 20, with the nodes colored by CONCOR block in the left hand column and edge-betweenness community in the right hand column. We notice that many of the communities identified by edge-betweenness, such as the orange community in the upper-left of the exam 1 network, has members from two blocks (lavender and grey), the lavender nodes are only connected to the rest of the network through the grey node. Other communities display similar properties, where there are a set of nodes that are more central to the community, and other nodes more peripheral to the community. This peripheral participation in the community is either due to that node being more strongly connected to other communities in the network (such as the grey node previously mentioned) or being more isolated from the community (such as the node at the top of the green community to the right of the orange community in the exam 1 network).

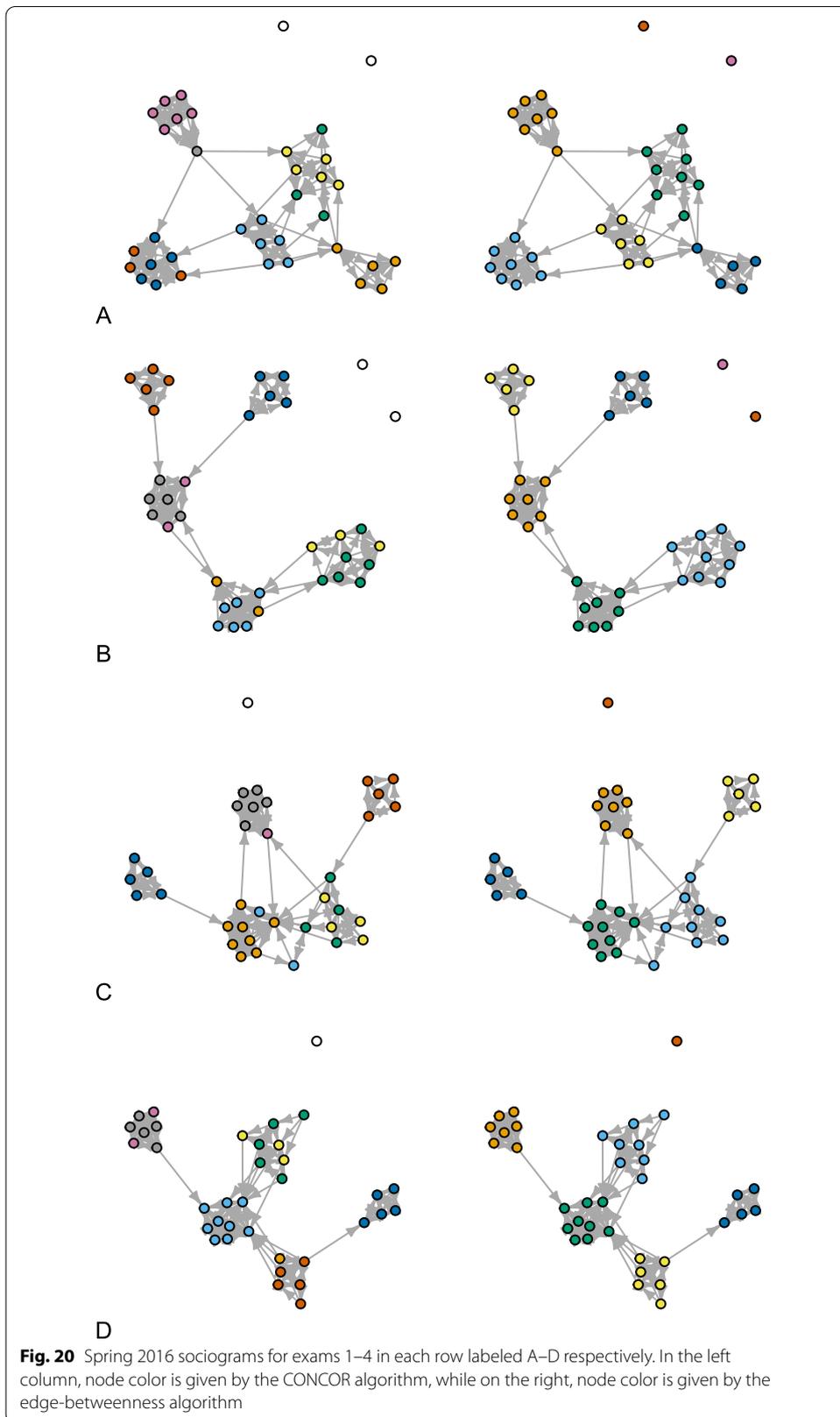


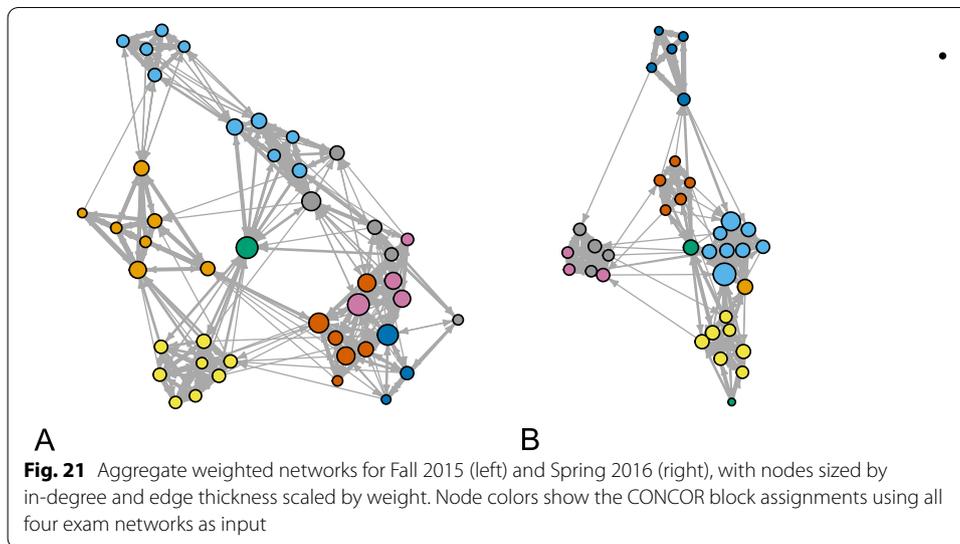
Aggregate CONCOR results

At the level of two CONCOR splits, essentially all the reduced networks look the same—they consist of “island” positions which connect internally with not enough exterior links to exceed the display threshold. This corresponds to a “coherent subgroups” structure, which has been observed in other active learning classrooms (Traxler et al. 2020). The three-split structure shown in Figs. 4 and 13 shows more complexity and numerous bridges between positions. Additional context that CONCOR can add, and which most community detection algorithms cannot, is a blocking for the four-exam sequence that uses each “snapshot” of links to group by linking behavior through the entire semester.

Figure 21 shows the weighted full-semester networks colored by this multi-time-point block assignment. A few patterns emerge in this view that are not visible at the single-exam level. In Fall 2015, one node with high in- and out-degree is distinct enough in linking behavior to form its own cluster (green); this person shifted through different blocks during the semester and did not follow the general trend toward “settling down.” Another block (yellow) was a coherent subgroup that largely stayed the same through the semester. Several smaller groups (red, purple, dark blue,







gray) are well connected to each other in aggregate, but split and reformed in various configurations over different exams.

In Spring 2016, the general stability of the network shows in a more modular structure in Fig. 21B, with fewer links between clusters than in the fall. Two small blocks (green and orange) consist of nodes that tend to be on the border of other clusters during the semester, appearing as a bridge point between more consistent groups. For most other blocks, the tendency is toward a high degree of internal communication and a less diverse set of bridging connections.

These time-sequenced CONCOR results, when compared with the blockings from individual exams, can identify students who form the core or nexus of a collaboration group, as distinct from others who are “short-term visitors” for one or two exams. From an instructor’s point of view, these nuances of collaboration are very difficult to capture in real-time, so the network results allow for a more thorough evaluation of how the group exam process played out. When combined with exam scores (the subject of ongoing analysis), this can also give a sense of how effective students’ self-directed groupings were at pooling their knowledge for the exam.

Study limitations

One of the limitations of this study is that, while group work is commonplace in schools at all levels, group assessment—in particular, high stakes group assessment—is much less common. Changes in the network could indicate growing familiarity with the group assessment paradigm rather than how the classroom social network is actually changing.

Finally, many studies in SNA look at multi-relational data. We did not collect any data in this regard due to the restrictions in our IRB. Students are indeed relating to one another in multiple ways that we are not capturing with these exam networks. For example, students who take these physics courses often take calculus courses concurrently, and have opportunities to interact in that setting.

Conclusions

We observed self-reported student collaboration networks on group exams for students working in a two-semester (Fall/Spring) introductory physics course. Classrooms are spaces where social relationships are important to learning, and networks give us a set of tools for understanding classroom social relationships. Previous work suggested that before each exam was graded, students in exam networks were able to identify higher-scoring peers on exams late in each semester, but not early in the semester (Wolf et al. 2017). We were interested in better understanding these networks from a centrality perspective as well as from a blockmodeling perspective using CONCOR (Breiger et al. 1975). We found that the fall semester was a “feeling out period” where students developed ties. Nodes in these networks began to develop more stable centrality properties, and close to the end of the fall semester, block membership became more stable as well. In the transition from the fall to the spring semester, about half of the class remained the same, and were able to leverage existing relationships to allow the class to build familiarity with each other more quickly. Centrality scores established on the first exam did not last to the second exam, but centrality scores established on the second exam tended to last at least until the next exam. CONCOR block membership established on the first exam remained stable throughout the spring semester, suggesting that the core group from the fall semester were able to provide enough structure to the class to facilitate this social development.

In comparing the CONCOR and edge betweenness network partitions, we find that they illuminate related but non-identical facets of the social structure. Edge betweenness is superior for finding closely-connected subgroups and is not limited to assigning 2nd partitions. CONCOR can distinguish different linking behaviors that may have social or educational significance, such as identifying students who are marginal to a more densely-connected community. We recommend comparing both sets of partitions to see a more complete picture of network structure.

In the future, we plan to expand our study to include students’ race, gender, and major. One of the things that we try to pay attention to is identifying at-risk students. Often, institutions focus on demographic factors such as if a student is a first-generation student. Understanding how social position within a classroom predicts performance in the current course, performance in future courses, and retention to the major is of vital importance to institutions where factors like these continue to come under more scrutiny.

Abbreviations

PER: Physics Education Research; ECU: East Carolina University; SFW: Steven F. Wolf; TMS: Timothy M. Sault; CONCOR: Convergence of iterated Correlations; AvgCC: Average local clustering coefficient; AvgDist: Average vertex–vertex distance ignoring disconnected node pairs; AvgDistUC: Average vertex–vertex distance including disconnected node pairs; TS: Tyme Suda; ALT: Adrienne L. Traxler.

Acknowledgements

SFW would like to thank Hunter Close, who got him started on group exams way back in 2013, and introduced him to some of the group exam literature in PER. SFW would also like to thank Cody Blakeney who helped design a GUI to aid in the entry of the network collaboration data. This study was approved by the East Carolina University IRB (UMCIRB 17-002031).

Author contributions

SFW taught the classes that are being studied here and supervised TMS. TMS did the initial coding on this project especially cleaning network data. TS is the main coder on the `concorR` package. ALT supervised TS work on the `concorR` package. ALT and SFW have been working together to characterize these networks coordinating the CONCOR results

with the previous work, and have also developed additional functionality for the `concoR` package. All authors read and approved the final manuscript.

Funding

TS was supported by NSF DUE-1712341 in work that developed the `concoR` package used in this analysis.

Availability of data and materials

Data will not be made publicly available per IRB restrictions.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 10 January 2022 Accepted: 17 April 2022

Published online: 04 May 2022

References

- Beatty I (2015) Collaboration or copying? Student behavior during two-phase exams with individual and team phases. In: Physics education research conference 2015. PER conference, College Park, MD, pp 59–62. <https://doi.org/10.1119/perc.2015.pr.010>
- Biancani S, McFarland DA (2013) Social networks research in higher education. In: Higher education: handbook of theory and research. Springer, Dordrecht, pp 151–215
- Bodin M (2012) Mapping university students' epistemic framing of computational physics using network analysis. *Phys Rev Spec Top Phys Educ Res* 8(1):010115. <https://doi.org/10.1103/PhysRevSTPER.8.010115>
- Bonacich P (1987) Power and centrality: a family of measures. *Am J Sociol* 92(5):1170–1182
- Breiger RL, Boorman SA, Arabie P (1975) An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling. *J Math Psychol* 12(3):328–383. [https://doi.org/10.1016/0022-2496\(75\)90028-0](https://doi.org/10.1016/0022-2496(75)90028-0)
- Brewe E, Kramer LH, O'Brien GE (2010) Changing participation through formation of student learning communities. In: AIP conference proceedings, vol 1289, pp 85–88. <https://doi.org/10.1063/1.3515255>
- Brewe E, Kramer L, Sawtelle V (2012) Investigating student communities with network analysis of interactions in a physics learning center. *Phys Rev Spec Top Phys Educ Res* 8(1):010101. <https://doi.org/10.1103/PhysRevSTPER.8.010101>
- Brewe E, Bruun J, Bearden IG (2016) Using module analysis for multiple choice responses: a new method applied to force concept inventory data. *Phys Rev ST Phys Educ Res* 12(2):020131. <https://doi.org/10.1103/PhysRevPhysEducRes.12.020131>
- Bruun J (2016) Networks as integrated in research methodologies in PER. In: Physics education research conference 2016. PER conference plenary paper. Sacramento, CA, pp 11–17
- Bruun J, Bearden IG (2014) Time development in the early history of social networks: link stabilization, group dynamics, and segregation. *PLoS ONE* 9(11):112775. <https://doi.org/10.1371/journal.pone.0112775>
- Bruun J, Brewe E (2013) Talking and learning physics: predicting future grades from network measures and Force Concept Inventory pretest scores. *Phys Rev Spec Top Phys Educ Res* 9(2):020109. <https://doi.org/10.1103/PhysRevSTPER.9.020109>
- Carr ET, Sault TM, Wolf S (2018) Student expectations, classroom community, and values reported on group exams. In: Traxler A, Cao Y, Wolf S (eds) Physics education research conference 2018. PER conference, Washington, DC
- Cochran GL, Gupta A, Hyater-Adams S, Knaub AV, Roman BZ (2019) Emerging reflections from the people of color (POC) at PERC discussion space. [arXiv:1907.01655](https://arxiv.org/abs/1907.01655)
- Commeford K, Brewe E, Traxler A (2021) Characterizing active learning environments in physics using network analysis and classroom observations. *Phys Rev Phys Educ Res* 17:020136. <https://doi.org/10.1103/PhysRevPhysEducRes.17.020136>
- Csardi G, Nepusz T (2006) The igraph software package for complex network research. *Inter J Complex Syst* 1695:1–9
- Dawson S (2008) A study of the relationship between student social networks and sense of community. *Educ Technol Soc* 11(3):224–238
- Ferligoj A, Doreian P, Batagelj V (2011) Positions and roles. In: Scott J, Carrington PJ (eds) The SAGE handbook of social network analysis. SAGE, Thousand Oaks, pp 434–446
- Forsman J, Linder C, Moll R, Fraser D, Andersson S (2014) A new approach to modelling student retention through an application of complexity thinking. *Stud High Educ* 39(1):68–86. <https://doi.org/10.1080/03075079.2011.643298>
- Freeman LC (1978) Centrality in social networks conceptual clarification. *Soc Netw* 1(3):215–239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
- Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP (2014) Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci* 111(23):8410–8415
- Gilley BH, Clarkston B (2014) Collaborative testing: evidence of learning in a controlled in-class study of undergraduate students. *J Coll Sci Teach* 43(3):83–91
- Ginsburg HP, Opper S (1988) Piaget's theory of intellectual development, 3rd edn. Prentice-Hall Inc, Englewood Cliffs
- Hake RR (1998) Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. *Am J Phys* 66(1):64–74
- Hirtle JSP (1996) Social constructivism. *Engl J* 85(1):91
- Institutional Planning, Assessment, and Research (2016) Fact book 2015–2016. Technical report, East Carolina University. <https://ipar.ecu.edu/wp-content/pv-uploads/sites/130/2020/01/Fact-Book-15-16.pdf>

- Koponen IT, Nousiainen M (2018) Concept networks of students' knowledge of relationships between physics concepts: finding key concepts and their epistemic support. *Appl Netw Sci* 3(1):1–21. <https://doi.org/10.1007/s41109-018-0072-5>
- Koponen IT, Pehkonen M (2010) Coherent knowledge structures of physics represented as concept networks in teacher education. *Sci Educ* 19(3):259–282. <https://doi.org/10.1007/s11191-009-9200-z>
- Leight H, Saunders C, Calkins R, Withers M (2012) Collaborative testing improves performance but not content retention in a large-enrollment introductory biology class. *Cell Biol Educ* 11(4):392–401. <https://doi.org/10.1187/cbe.12-04-0048>
- Lin Y, Brookes D (2013) Using collaborative group exams to investigate students' ability to learn. *AIP Conf Proc* 1513(1):254–257
- Luo W, Yin P, Di Q, Hardisty F, MacEachren AM (2014) A geovisual analytic approach to understanding geo-social relationships in the international trade network. *PLoS ONE* 9(2):88666. <https://doi.org/10.1371/journal.pone.0088666>
- Mastascusa EJ, Snyder WJ, Hoyt BS (2011) Effective instruction for STEM disciplines: from learning theory to college teaching. Jossey-Bass, San Francisco
- McCullough L, Chessey MK, Cochran GL, Cunningham B, Johnson A, Singh C (2019) Gender bias in physics: an international forum. *AIP Conf Proc* 2109(1):030007. <https://doi.org/10.1063/1.5110069>
- McDermott LC, Redish EF (1999) Resource letter: PER-1: physics education research. *Am J Phys* 67(9):755–767
- McDermott LC, Shaffer PS (2002) Tutorials in introductory physics. Prentice Hall, Upper Saddle River
- Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45(2):167–256. <https://doi.org/10.1137/S003614450342480>
- Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E* 69:026113. <https://doi.org/10.1103/PhysRevE.69.026113>
- Pons P, Latapy M (2005) Computing communities in large networks using random walks. In: Yolum P, Güngör T, Gürgeç F, Özturan C (eds) *Computer and information sciences—ISCIS 2005*. Springer, Berlin, pp 284–293
- Prell C (2012) *Social network analysis: history, theory & methodology*. SAGE, Thousand Oaks
- R Core Team (2020) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rosvall M, Bergstrom CT (2010) Mapping change in large networks. *PLoS ONE* 5(1):8694. <https://doi.org/10.1371/journal.pone.0008694>
- Sagr M, López-Pernas S (2022) The curious case of centrality measures: a large-scale empirical investigation. *J Learn Anal* 9(1):13–31. <https://doi.org/10.18608/jla.2022.7415>
- Sault TM, Close H, Wolf S (2018) Student cognition in physics group exams. In: *Physics education research conference 2018*. PER conference, Washington, DC
- Scott TF, Schumayer D (2018) Central distractors in force concept inventory data. *Phys Rev ST Phys Educ Res* 14:010106. <https://doi.org/10.1103/PhysRevPhysEducRes.14.010106>
- Suda T, Wolf SF, Traxler A (2020) *concorR: CONCOR and supplemental functions*. R package version 0.2.2. <https://github.com/ATraxLab/concorR>
- Traxler A (2015) Community structure in introductory physics course networks. In: Churukian AD, Jones DL, Ding L (eds) *2015 physics education research conference*, College Park, MD, pp 331–334. <https://doi.org/10.1119/perc.2015.pr.078>
- Traxler A, Gavrin A, Lindell R (2018) Networks identify productive forum discussions. *Phys Rev Phys Educ Res* 14(2):020107. <https://doi.org/10.1103/PhysRevPhysEducRes.14.020107>
- Traxler AL, Suda T, Brewé E, Commeford K (2020) Network positions in active learning environments in physics. *Phys Rev Phys Educ Res* 16:020129. <https://doi.org/10.1103/PhysRevPhysEducRes.16.020129>
- Vargas DL, Bridgeman AM, Schmidt DR, Kohl PB, Wilcox BR, Carr LD (2018) Correlation between student collaboration network centrality and academic performance. *Phys Rev Phys Educ Res* 14(2):020112. <https://doi.org/10.1103/PhysRevPhysEducRes.14.020112>
- Von Korff J, Archibeque B, Gomez KA, Heckendorf T, McKagan SB, Sayre EC, Schenk EW, Shepherd C, Sorell L (2016) Secondary analysis of teaching methods in introductory physics: a 50 k-student study. *Am J Phys* 84(12):969–974
- Wasserman S, Faust K et al (1994) *Social network analysis: methods and applications*, vol 8. Cambridge University Press, Cambridge
- Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393(6684):440–442. <https://doi.org/10.1038/30918>
- Wells J, Henderson R, Stewart J, Stewart G, Yang J, Traxler A (2019) Exploring the structure of misconceptions in the Force Concept Inventory with modified module analysis. *Phys Rev ST Phys Educ Res* 15:020122. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020122>
- Wieman CE, Rieger GW, Heiner CE (2014) Physics exams that promote collaborative learning. *Phys Teach*. <https://doi.org/10.1119/1.4849159>
- Wolf SF, Dougherty DP, Kortemeyer G (2012a) Empirical approach to interpreting card-sorting data. *Phys Rev ST Phys Educ Res* 8:010124. <https://doi.org/10.1103/PhysRevSTPER.8.010124>
- Wolf SF, Dougherty DP, Kortemeyer G (2012b) Rigging the deck: selecting good problems for expert-novice card-sorting experiments. *Phys Rev ST Phys Educ Res* 8:020116. <https://doi.org/10.1103/PhysRevSTPER.8.020116>
- Wolf SF, Doughty L, Irving PW, Sayre E, Caballero MD (2014) Just math: a new epistemic frame. In: *Physics education research conference 2014*. PER conference, Minneapolis, MN, pp 275–278
- Wolf S, Blakeney C, Close H (2016) Group formation on physics exams. In: Jones DL, Ding L, Traxler A (eds) *Physics education research conference 2016*. PER conference, Sacramento, CA, pp 400–403
- Wolf S, Sault TM, Close H (2017) Information flow in group exams. In: Ding L, Traxler AL, Cao Y (eds) *Physics education research conference 2017*. PER conference, Cincinnati, OH, pp 444–447
- Zwolak JP, Zwolak M, Brewé E (2018) Educational commitment and social networking: the power of informal networks. *Phys Rev Phys Educ Res* 14(1):010131. <https://doi.org/10.1103/PhysRevPhysEducRes.14.010131>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.