122nd ASEE Annual Conference & Exposition
June 14 - 17, 2015
Seattle, WA
Making Value for Society
Paper ID #12840

Using Havel-Hakimi to graph classroom networks

Dr. Jonathan C. Hilpert, Georgia Southern University

Jonathan C. Hilpert is an Educational Psychologist at Georgia Southern University.

Ms. Rebecca Holliday , Georgia Southern University

Bachelor of Science in Applied Mathematics from Middle Georgia State College. Currently a graduate student in the Department of Mathematical Sciences at Georgia Southern University with a concentration in Applied Mathematics and research in Graph Theory.

Using Havel-Hakimi to graph classroom networks

Engineering education researchers emphasize the importance of teamwork and collaborative learning (Borrego, et al., 2013). Researchers have argued that collaborations can be studied as emergent systems, where the actions of individual agents produce an outcome that is not reducible to its individual parts (Sawyer, 2001). The underlying structure of an emergent system is a network (i.e. Davis and Sumara, 2006; Watts, 1999). Networks can be graphically represented as interconnected collections of nodes and edges (Gross and Yellen, 2006), or in the context of the current study students working together in classrooms to learn and solve problems. Although researchers have examined collaborative emergent systems in classrooms from a qualitative perspective (Sawyer, 2004), there is need for more advanced tools to examine how the network structure of classrooms influences student learning and performance.

In this paper, we use the Havel-Hakimi algorithm to investigate classroom network graphs. The Havel-Hakimi algorithm is a popular method for determining whether a given degree sequence is graphical and is computationally inexpensive. The algorithm uses a recursive method to create a simple graph from a graphical degree sequence. In this case, the degree sequence is comprised of each student in a classroom and the number of peers they self report collaborating with during classroom activities. We expand upon the Havel-Hakimi algorithm by coding a program in Python that generates random graphs with the same degree sequence. In doing this, we can examine all potential possibilities of which students work with whom. Then, we use an edge-weight technique to select an isomorphic class of the classroom network. We analyze why the classroom network looks this way and what it means. To do this, we use Gephi (popular network analysis software) to calculate closeness, betweenness, and other measures of network characteristics.

Our results describe a useful technique for developing classroom graphs that can graphically represent engineering classroom networks. We show some example graphs and conclude with a discussion of how these graphs may be related to student learning.

Gathered Information

The data was gathered from engineering students attending a research university in the southeast. Students from two classrooms participated in the survey: a smaller upper level course (N = 12) and a large lower level course (N = 52). Over 80 percent of the students in each course completed the survey. Course titles and numbers are held back to protect the identities of the instructors. Part of the survey asked students to respond to questions about the type of instruction the professor uses and their levels of collaboration and cognition (defined below).

Students in the large lower level course reported hearing lectures over 70 percent of the time. Students in the smaller upper level course reported hearing lectures roughly 40 percent of the time. After answering questions about their classroom, the students were asked to report the number of students they typically work with during activities in class. Students responded by reporting a number, which we found ranged from 0 to 5 (n = 55). We gathered up the numbers for all student respondents and treated them as a *degree sequence*.

In graph theory, an area of discrete mathematics focused on the study of graphs, a degree sequence is defined as follows:

Definition. *For a graph G with vertices $\{v_1, \cdots, v_n\}$ with degrees $d_i = deg(v_i)$. Define the degree sequence of G to be the non-increasing sequence $\{d_{i_1}, \cdots, d_{i_n}\}$.*

For example, Figure 1 shows a graph whose degrees sequence is [2,2,2,1,1]. The graph contains 5 vertices. The number of edges, $|E|$, can be calculated,

$$\sum deg(v_i) = 2|E|.$$

In the above sequence, there are $\frac{2+2+2+1+1}{2} = \frac{8}{2} = 4$ edges. Graphs of this nature can be used to represent a range of social and natural phenomena including the world wide web, food chains, and the famous "small world" problem (see Strogatz, 2001 for a review). Here, we use them to represent classrooms.

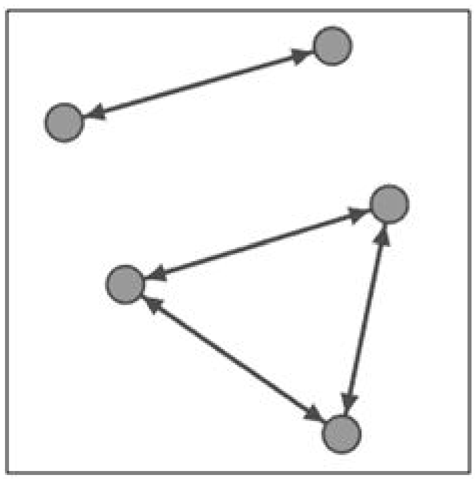

Figure 1: Graph with degree sequence [2,2,2,1,1]

Classroom Networks

A degree sequence for a classroom is a list individual students and the number of peers with which a student studied or collaborated to determine the degree. Consider the degree sequence $\{n_{1_1}, n_{1_2}, \cdots, n_{1_i}, n_{2_1}, n_{2_2}, \cdots, n_{2_j}, n_k, \cdots, n_{k_l}\}$, where n_{p_x} and n_{p_y} represents students who have the same ingroup number. Theoretically, all classroom networks are graphical if a perfect representation of the network is attained, but incomplete classroom network data may not be graphical by Havel-Hakimi requirements. This can be caused by students who have reported the wrong number of ingroup peers or by not having all students in the class take the survey, or some combination of the above.

Thus, although we had large samples from each of the test courses (i.e. greater than 80 percent), we needed to examine the trustworthiness of the data. We began by examining the degrees sequence data to determine if any outliers existed in the data (i.e. students who reported questionable numbers of partners). No outliers were found in the data. Next, we used the Havel-Hakimi Theorem to determine if the degree sequences were graphical. Raw data from both classrooms proved graphical with no modification. The Havel-Hakimi algorithm we used to determine if a simple graph existed for the degree sequence data was based on the following theorem:

Theorem (Havel (1955) Hakimi (1962))**.** *Let $S = (d_1, \cdots, d_n)$ be a finite list of nonnegative integers that is nonincreasing. List S is graphic if and only if the finite list $S' = (d_2 - 1, d_3 - 1, \cdots, d_{d_1+1} - 1, d_{d_1+2}, \cdots d_n)$ has nonnegative integers and is graphic.*

The algorithm then follows a recursive method where the theorem is used with $S = S'$ until $S' = \{0\}$.

Generating Random Graphs

Once we determined the classroom networks in question were graphical, we wanted to generate a random graph from the degree sequence that was likely to represent the network structure of the data. The inspiration came from a popular method in statistics called bootstrapping. Using Havel-Hakimi Algorithm, a program was written in python and run in MATLAB that could generate a single random graph from a graphical degree sequence. Expanding on this, an additional program was written to perform "edge swapping." To do this, We borrowed code from other researchers interested in similar problems (found at website `http://strategic.mit.edu/downloads.php?page=matlab_networks`).

Our expanded program is able to swap edges within a random graph, generating as many random graphs as desired from one graphical degree sequence. For each classroom degree sequence, we ran the algorithm 100 times to generate 100 random graphs. It should be noted that it is possible for the same graph to be generated more than once. We asked the program to count the number of times an edge appeared in all 100 graphs and developed an edge frequency list. Then, for each individual graph, we summed the edge frequencies. The graph with highest sum became our best fit model because it represents the graph with the most number of edges common across isomorphic classes.

After running the algorithm on 100 iterations, we used Gephi, a popular network analysis program, to visualize the results. We used the Forced Atlas II layout algorithm to examine the graphs for evidence of face validity and to determine what the network indicated about the groups that formed. Because we knew, a priori, that one was a large lecture course the other was a smaller laboratory course, and we also knew what types of instructional strategies were being used in each from student self reports, we expected to see patterns in the graph that might be representative of networks related to these instructional styles.

Additionally, Gephi has the capability of analyzing network structures and producing statistics that describe individual nodes and network characteristics. We choose to examine the number of modularity classes in the networks, and measures of closeness centrality and betweenness centrality for the individual nodes. The modularity algorithm identifies the number of likely ”communities” in a graph, in this case the number of working groups. Betweeness centrality measures how often a node appears on shortest paths between nodes in the network. Closeness Centrality measures the avergage distance from a given starting node to all other nodes in the network. In this case, both are measures of information sharing among classmates.

Results

Our inspection of the graph layout provided face validity evidence for the network structure of the classroom. The small, upper division course formed a network with four modularity classes, or communities of students. The large lower division course, on the other hand, formed a network with 36 modularity classes, 32 of which were comprised of single students (students with zero degrees were treated as single modularity classes). See figure 2 for the graphs that were generated from the data, with modularity classes represented in gray scale. Thickness of edge weight represents edge frequency from random simulation.

Student closeness scores were significantly and positively correlated with self reported collaboration ($r = .33$; $p < .01$, $n = 55$) and cognition ($r = .20$; $p < .05$; $n = 55$). Student betweenness measures were significantly and positively correlated with self reported collaboration ($r = .27$; $p < .01$, $n = 55$) but not cognition ($r = .15$; $p = .08$; $n = 55$). Collaboration was assessed with a self reported Likert response (7 point agree or disagree scale) to the item, ”My classmates and I actively worked together to complete assignments.” Cognition was assessed using a self report Likert response (7 point agree or disagree scale) to the item, ”I thought about different approaches or strategies I could use for studying the assignments.”

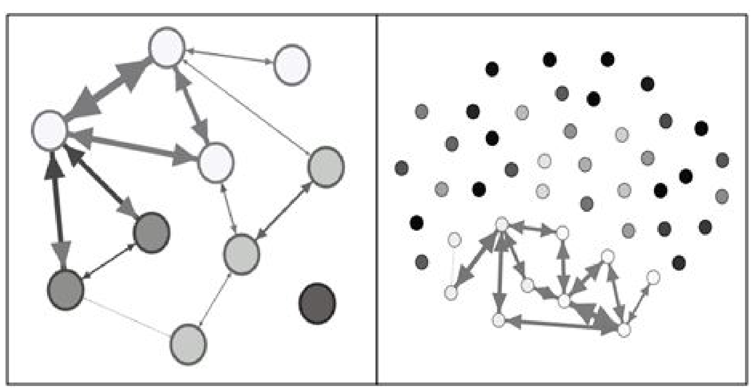

Figure 2: Generated graphs for small course and large course respectively

Discussion

Classroom network analysis may be a useful way of examining collaboration in engineering education classrooms. Network layouts and measures such as closeness centrality and betweenness centrality statistics have the potential to provide important information about impact of instruction at the group level of analysis. Our pilot comparison suggested that in the large lecture course we surveyed, a tight network – with the same number of modulartity classes as the small upper division course – emerged in the classroom. The face validity of these networks was supported with our statistical evidence that students embedded in these networks were more likely to collaborate (evidence of concurrent validity) and strategically learn (evidence of predictive validity).

These findings are impactful because of the large number of students who were not embedded in the classroom network. Our preliminary evidence suggests these students may be less likely to consider different approaches or strategies for learning, though some of these students may be high performing and prefer not to collaborate. Perhaps encouraging networks with multiple diads or triads, or collections of distinct modularity classes, in larger courses will engage a larger percentage of the students who might benefit from collaboration. It is not clear what types of instructional strategies might be related to different network structures, or how much variation in student learning classroom networks might predict, but these small scale findings suggest it might be worth examining further using survey research.

Our findings also suggest that expanding the Havel Hakimi approach to include an edge swapping technique is an effective method for generating random graphs from student degree sequences. One potential limitation here, that we continue to explore, is variation in isomorphic class. Our edge swapping technique creates a random graph, choosen by statistical probability. Thus, the isomorphic class of the best fitting graph is chosen at random, and it is not clear if small differences in isomorphic classes have theoretical implications for classroom instruction. Variation in isomorphic classes for large graphs continues to be a challenge for mathematicians, and we plan to continue to run monte carlo studies to examine the problem from a educational, theoretical perspective.

Network analysis has provided new insights into a wide variety of social phenomena, particularly with regard to understanding the underlying structures of complex systems that produce emergent outcomes. Classrooms have been described in the extant literature as complex adaptive systems (Sawyer, 2004), and the emergent nature of educational environs has been established (Davis and Sumara, 2006). The results of our pilot study, presented here, suggest there may be value in systematically studying classrooms using network analysis to examine the relationship between 1) instructional categories and network structure, and 2) network and structure and student cognition for mediational relationships.

References

Borrego, M., Karlin, J., McNair, L. D., and Beddoes, K. (2013). Team effectiveness theory from industrial and organization psychology applied to engineering student project teams: A research review. Journal of Engineering Education. 102(4), 472-512.

Davis, B. and Sumara, D. (2006). Complexity and Education: Inquires Into Learning, Teaching, and Research. Lawrence Erlbaum and Associates, Mahwah, NJ.

Gross, J. and Yellen, J. (2006). Graph Theory and its Applications. Taylor and Francis. Boca Raton, Fl.

Hakimi, S. L. (1962). On realizability of a set of integers as degrees of the vertices of a linear graph. Journal of the Society for Industrial and Applied Mathematics 10: 496506, MR 0148049.

Havel, V. (1955) A remark on the existence of finite graphs. asopis pro pstovn matematiky (in Czech) 80: 477480.

Sawyer, R. K. (2001). Emergence in Sociology: Contemporary Philosophy of Mind and Some Implications for Sociological Theory American Journal of Sociology, Vol. 107, No. 3, pp. 551-585.

Sawyer, R. K. (2004). Improvised Lessons: Collaborative Discussion in the Constructivist Classroom. Teaching Education. 15(2), 189-201.

Watts, D. J. (1999). Small Worlds: The Dynamics of Networks Between Order and Randomness. Princeton University Press.