Student-Course Network Analysis Proposal

Kelly Gothard

UVM Office of Institutional Research

**Introduction**

Student engagement in community-based activities is a valuable indicator for a student's future success. The specific tasks associated with these activities can stimulate a student's interest in their major or favorite recreational activities, but the social benefits that come with being engages in community-based activities are arguably just as important in maintaining and strengthening a student’s engagement with the university. The most fundamental source of socialization in universities is in the classroom, where students form bonds through asking questions, working on homework, and group activities. Tightly knit communities of students that take the same courses may be more likely to study together, engage in extracurricular activities together, encourage class attendance, and also encourage retention. The connectedness of students within their courses may be an indicator for future success at the university, and can be studied by looking at the network of students to courses.

**Dataset**

The dataset will contain an entry for each course that a student took, the student’s information (identifier, major, college, etc), for some number of semesters. The privacy of student data is maintained through group-wide analysis. Information on each student is necessary to construct the network, because each individual node needs to be accounted for. However, our interest is in the patterns that emerge from the network, such as number of connections per node. The results from this study will not pertain to an individual student.

We cannot aggregate student rows to de-individualize the dataset because each student has a unique course sequence, and we need to map students to the courses that they took at a particular time and associate them with other students via mutual courses taken. The observations made during this study will be regarding communities of un-identifiable students and distributions of student connectivity.

**Methods**

A bipartite network will be constructed, where nodes on one side represent students and nodes on the other side represent courses and an edge between nodes represents a student that took a course (see Figure 1 below for an example). The projection of the bipartite network onto the ‘student space’ will show each student in the dataset. Connections of weight n between students indicate that a student took n classes with another student (see Figure 2). This network is a useful representation of the social network that is created by students taking classes together, and the position of a student on that network can be used to determine a measure of connectivity for an individual student.

Social networks often exhibit homophily-driven bias; birds of a feather flock together. In this project, students who are in the same major program will already be taking courses together, and some major programs are more ‘tight-knit’ than others. For example, the electrical engineering program has a strict set of courses and a strict order that those courses should be taken in, whereas the data science major is far less structured and has various paths and electives that could be taken by data science students. To account for variation among major programs, the network will be ‘chopped up’ into subnetworks of each major, and each student’s relative connectivity will be measured in their major group. This process can also be repeated by college.

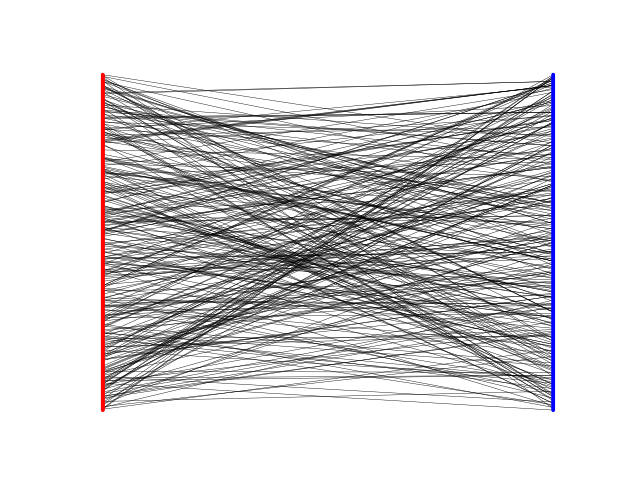
We will then study the relationship between course-facilitated social connectivity and GPA, retention, major-changes, college-changes, and more. We can also measure inter-major connectivity, inter-college connectivity, inter-race/ethnicity connectivity, etc to quantify the level of interaction between various groups at UVM.

**Software**

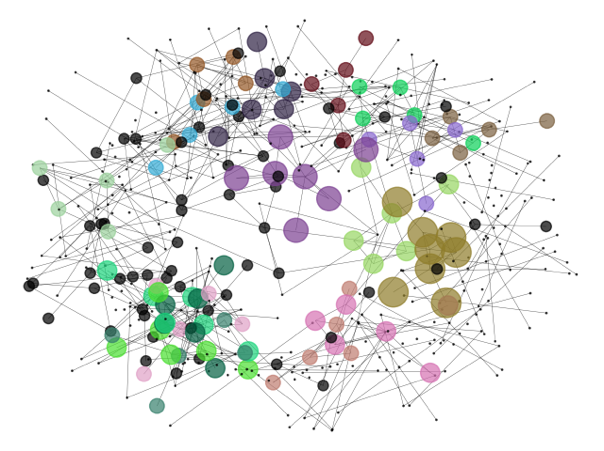
To construct a network data structure, detect communities, create visualizations, and compute measures of connectivity, programs will be written in Python 3.7.1. We anticipate the use of open-source Python libraries found in Table 2, and such dependencies will be included in a transportable Python environment upon conclusion of this project so that it may be used in the future.

Large networks, such as the student-course connectivity network that we will be working with in this study, can require supplemental computing resources to maintain a reasonable timeline of computational processes. For high-volume computational tasks such as community detection, we will utilize the Vermont Advanced Computing Core.

**Figures**



**Figure 1. Bipartite Network of Students and Courses from Artificially Generated Dataset:** On the left, each student is represented by a red dot, or node. There are many students here (1000), so the dots blend together to look like a line. The courses that these students take are represented as blue nodes on the right, and a connection, or an edge, is drawn between a student and a course if a student has taken that course.



**Figure 2. Projection of Bipartite Network onto Student-Space**: Each node represents a student, and an edge between nodes means that those students took a class together. This network was created by taking the projection of the bipartite network in Figure 1, and connecting students that share connections to courses. Nodes are colored based on detected community, and nodes are sized based on the size of the community, raised to the power of three for visualization purposes.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Individual Identifier (ID) | College1 | College2 | Major1 | Major2 | Course | Semester |
| Student1 | CEMS | CEMS | DS | DS | CS 021 | F 2018 |
| Student1 | CEMS | CEMS | DS | DS | MATH 121 | S 2019 |
| Student2 | CALS | CAS | CDAE | PSYS | PSYS 001 | F 2018 |

**Table 1. Dataset Schema Example**

**Bibliography**

1. Barber, M. J. (2007). Modularity and community detection in bipartite networks. Physical Review E, 76(6), 066102
2. Cunha, T. O., Weber, I., Haddadi, H., Pappa, G. L. (2016, April). The effect of social feedback in a reddit weight loss community. InProceedings of the 6th International Conference on Digital Health Conference (pp. 99-103). ACM.
3. Fire, M., Guestrin, C. (2016). Analyzing complex network user arrival patterns and their effect on network topologies. arXiv preprintarXiv:1603.07445.
4. Heck, R. H., Price, C. L., & Thomas, S. L. (2004). Tracks as emergent structures: A network analysis of student differentiation in a high school. American Journal of Education, 110(4), 321-353.
5. Rosvall, M., Bergstrom, C. T. (2010). Mapping change in large networks. PLoS One, 5(1) doi:http://dx.doi.org/10.1371/journal.pone.0008694
6. Holme, Petter, and Jari Saramki. ”Temporal networks.” Physics reports 519.3 (2012): 97-125.
7. Diebold, F. X., & Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. Journal of Econometrics, 182(1), 119-134.
8. Kardan, A. A., Sadeghi, H., Ghidary, S. S., & Sani, M. R. F. (2013). Prediction of student course selection in online higher education institutes using neural network. Computers & Education, 65, 1-11.
9. Valente, T. W., & Foreman, R. K. (1998). Integration and radiality: Measuring the extent of an individual's connectedness and reachability in a network. Social networks, 20(1), 89-105.
10. White, D. R., & Harary, F. (2001). The cohesiveness of blocks in social networks: Node connectivity and conditional density. *Sociological Methodology*, *31*(1), 305-359.