James Max Kanter
Faculty Supervisor: Una-May O'Reilly, Kalyan Veeramachaneni
Term: Spring 2014

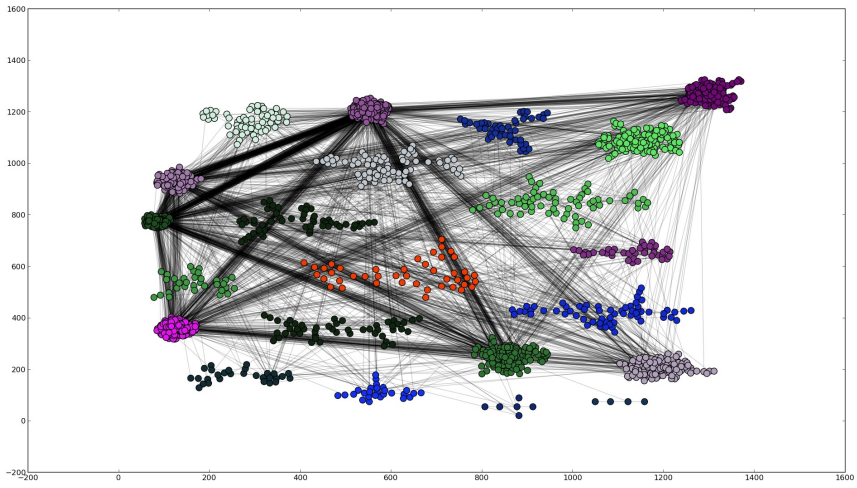
Proposal : Analyzing Communities that Form in MOOC Discussion Forums

Project Overview
Massively online open courses, or MOOCs, are being offered by universities across the country, including MIT. Over 100 classes are offered on platforms like edX and Coursera and this number is growing. Due to the nature of being online and open, a lot of data is being collected about how students interact with the course material. In this project, I will focus on data collected from students participating in the discussion forums. This data will be used to analyze characteristics of past courses and help inform decisions about future courses.

Previous Work
Due to the short time MOOCs have been available, not many previous studies have been done to understand how students interact with a course.

Previous work such as has focused on features of individual users rather than looking at the interactions between users (Kizilcec et al, 2013).

During IAP 2014, I began investigating the potential to include features of interpersonal interactions to understand and characterize MOOCs. This investigation involved looking at one class and creating a graph of students who interacted with each other. I applied some preliminary clustering and was able to generate the following image.



This picture motivates future investigation into the underlying community structure of how

students interact.

Research Goals
Qualitatively explaining the community structure in MOOCs and determining the implications for educators is the overall goal of my project.

I hope to be answer questions such as

1. How many communities of students are there in a class?
2. What characteristics of students determine what community they belong to?
3. Can association with a community be used to predict course completion?
4. How does the description of the communities in a course change over time?
5. How are humanities classes different than engineering classes?

Milestones and time estimates

There are two phases of this project. Time estimates can be convert to weeks by using an estimate of 6 hrs of work per week.

<u>Phase I – Data extraction</u>

1. Create reusable procedure for extracting relevant data into graph structure – 12 hrs
2. Investigate and select algorithm for community detection – 18 hrs
3. Determine which characteristics are important to extract from community structure – 18 hrs
4. Implement that ability to track community structure over time – 12 hrs

Phase I Total – 60 hrs

<u>Phase II – Data analysis</u>

1. Prioritize which of the proposed (or new) questions above are most important – 6 hrs
2. Write report – 30 hrs

Phase II Total – 36 hrs

Project Total – 96 hours

Personal Statement
I had an excellent experience working with the ALFA group over IAP and I am looking forward to continue working into this semester. I hope to be able to produce an M.Eng thesis proposal for this project at the end of the semester.

Bibliography

René F. Kizilcec, Chris Piech, and Emily Schneider. 2013. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK '13), Dan Suthers, Katrien Verbert, Erik Duval, and Xavier Ochoa (Eds.). ACM, New York, NY, USA, 170-179. DOI=10.1145/2460296.2460330 http://doi.acm.org/10.1145/2460296.2460330