Using a Proximity Network on a University Campus as Supplemental Tool for Backwards Contact Tracing

Bryn Loftness

Introduction

From August 2020 to May 2021, WiFi connectivity data was collected using an external company, Degree Analytics, for the Colorado Mesa Campus in all university-affiliated spaces on the main campus. This included academic buildings, residential halls, student dining locations, administrative buildings, and athletic spaces. Data was initially first tracked by Degree Analytics for each individual, and then parsed in pair-wise format using a proprietary algorithm to slim potential pairings to those more proximity relevant. The pairwise data was then shared with our group, which included the two individuals, the location(s) of simultaneous access, the duration of the access, and the time and date. Pair-wise data was only shared for simultaneous access of two individuals for periods of longer than 15 minutes. Due to the proprietary nature of the parsing and re-aggregation of the proximity data, we can not confirm if two individuals were within reasonable range of each other to cause transmission of COVID-19, as some access points span large spaces in heavily trafficked buildings. Students were also mandated to wear masks and social distance within all academic buildings and residential halls until a protocol change was enacted in early March of 2021 during second semester. On March 2nd, a protocol change was announced that allowed students to enter residential buildings other than their own. Reduced mask mandates were put in place for students that could provide vaccination status also around that time frame.

Though we can not with full confidence identify transmission events with this data, we can examine these interactions to infer potential social relationships between individuals. With the extensive metadata collected from Colorado Mesa University including test dates, symptom onset dates, sports team involvement, club involvement, class year, gender, on-campus residency information, and extensive data about residential hall structure, we can link users within the proximity network dataset to the user IDs of individuals who tested positive throughout the course of the Fall 2020 and Spring 2021 semester and conjecture social connections. For a portion of the cases, we can also link individuals within the proximity network to genomic sequencing data.

We investigate whether utilization of a proximity network provides an advantage in reconstructing the spread of COVID-19 on the Colorado Mesa University campus for the Fall 2019 and Spring 2020 school year. This analysis could lead to novel innovations for contact tracing technologies utilizing Wi-Fi based proximity networks as a tool for more accurate and objective close-contact identification in disease spread on university campuses.

Background

In the nine months of the Fall 2020 to Spring 2021 school year, there were approximately 75.7 million pairwise simultaneous-access connections recorded. During the same nine-month window, there were 1361 positive cases linked to students, faculty, and staff on the Colorado Mesa University campus. 890 of the 1361 positive cases are present in the Wi-Fi dataset. Students, faculty, and staff are offered an opt-out procedure for Wi-Fi tracking, and thus did not capture every student, faculty, and staff member on

who was on campus over the course of the school year. Case count over time for both semesters is shown in Figure 1 below.
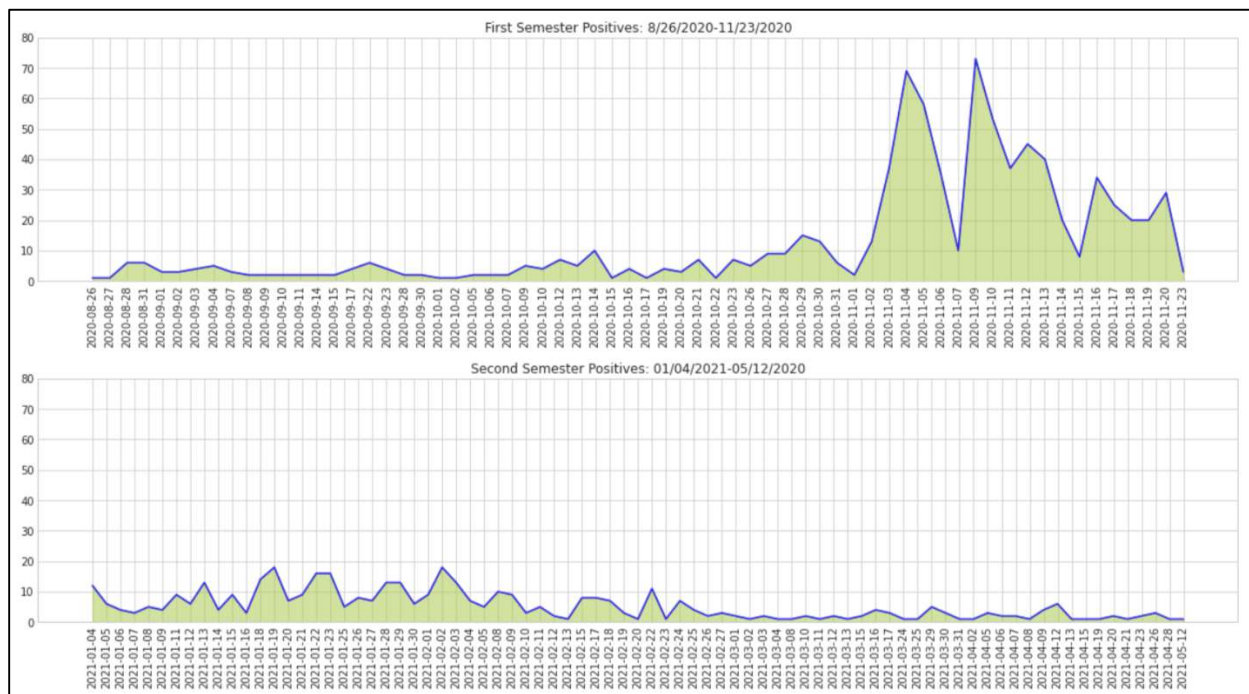

Figure 1 - Case Count Over Time For Fall 2020 and Spring 2021

Analysis using the proximity network was divided into three categories: Verifying expected on-campus dynamics; Examine validity of proximity network usage as an epidemiological tool; Using the proximity network as a guide to model the identification of potential close contacts for individuals who test positive within the network. Our analysis was primarily focused on categories one and two for the purposes of this examination.

Analysis:

Category 1 Proximity Network Analysis: *Showing Validity of the Network through Verification of Expected On-Campus Population Dynamics*

We first analyze daily and weekly patterns of movement and student interactions throughout the building types on campus. These dynamics can be visualized in Figure 2 below. We see by examining Fall 2020 data that there is a decrease in the number of interactions on Saturdays and Sundays within residential halls, while the interaction counts for residential halls stay relatively consistent throughout the week. There is some variability in the number of interactions throughout the week in academic buildings, which could be influenced by classroom schedule, building layout, or the increase in the option for hybrid classroom flexibility during Fall 2020 for academic courses. For example, more students could be electing to take their courses in the later half of the week online while the first half of the week they are in-person. We also see a decrease in Non-Residence and Non-Academic Halls during the weekends. These buildings include administrative buildings, dining halls, and athletic facilities. These findings are consistent with the dynamics we'd expect from on-campus students living and residing in the residential halls maintaining

consistent interaction counts throughout the week while the other locations on campus exhibit fluctuations based on weekday.
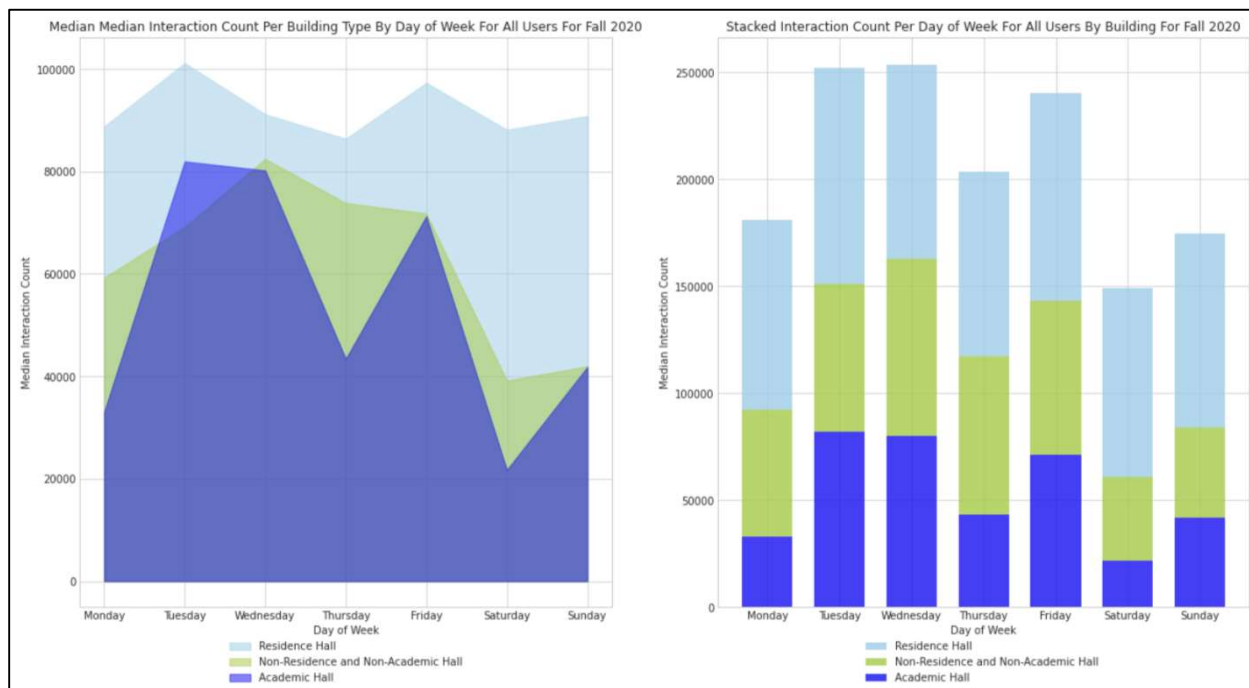

Figure 2 - By Day Interaction Count Per Building Type

We can also examine median duration time per interaction per day of week for academic versus residential buildings for the Fall 2020 semester. Again, we see relatively consistent duration lengths for Residential buildings throughout the week, while academic buildings have longer median interaction durations during the weekdays and peak median interaction durations on Tuesdays, Thursdays, and Fridays.

Under typical circumstances, the Fall 2020 semester would have ended 12/10/2020. However, due to the increase in positive cases on campus, CMU chose to transition to online-only for the weeks following Thanksgiving break until the end of the Fall 2020 semester. The official start date for the Spring 2021 semester was 1/25/2021, however some students started moving back to campus in early January for in-person J-Term courses. J-Term courses started 1/04/2021. The figures for Spring 2021 showing the same metrics of interaction count and median interaction duration per day are shown in the Supplemental Material section. The patterns seen in Spring 2021 do not reflect the dynamics we'd expect and see from Fall 2020, in that interaction levels are doubled while the number of unique individuals on campus each day are relatively the same as Fall 2020, shown in Figure 2 of the Supplemental Materials Section. We also see increased time spent in Residential
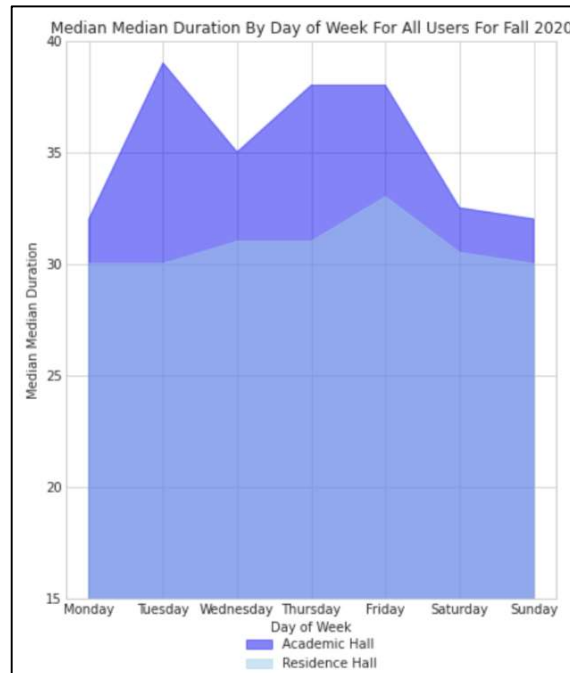

Figure 3 - Median Interaction Duration Median by Building

Halls and a decreased amount of time spent in Academic buildings, shown in Figure 1 of the Supplemental Materials section. We hypothesize this transition to be related to the relaxed protocols of students entering other residential buildings to visit friends, and the decreased restrictions put on movement and activity on campus. Based on these changes in campus dynamics, we focus our analysis primarily on Fall 2020 for these write-up.

Category 2 Proximity Network Analysis: *Showing Potential of the Network For Epidemiological Surveillance*

After examining that we could observe expected dynamics, we investigated the use of the Colorado Mesa University (CMU) proximity network as a tool for contact tracing. We test this hypothesis by evaluating the network attribute assortativity coefficient for each day of the semester. Each day, the students on campus change and thus the network is entirely new for each day of the semester. For this purpose, each day of the week a new network is constructed based on the individuals recorded in a pair-wise interaction for that day.

For each day, the undirected network is constructed and each node – corresponding to an individual in the proximity network – is given the attribute corresponding to their infection status over the course of the semester. If an individual at any point during Fall 2020 tests positive for COVID-19, that node is assigned the attribute of 1. If the individual does not test positive sometime during Fall 2020, the node is assigned the attribute of 0, regardless on if the individual tests positive during second semester. Once attributes are assigned to all nodes in the network, using the NetworkX package, numeric assortativity coefficient, the attribute assortativity of the network is recorded for that day.

Attribute assortativity quantifies the correlation to which connected nodes, i.e. nodes that share an edge and thus were recorded in a pairwise connection in our dataset, share the same attribute. The coefficient is measured from -1 to 1, where a coefficient of -1 represents the network being completely disassortative. This would mean that all nodes are connected to only nodes of the opposite or differing attribute. In our context, meaning that all individuals who eventually test positive at some point in Fall 2020 never have a pair-wise simultaneous access with another individual who eventually tests positive during Fall 2020. An attribute assortativity coefficient of 1 would imply the opposite, where pair-wise connections of Fall 2020 only exist with two individuals who eventually test positive in Fall 2020 or two individuals who never test positive in Fall 2020. Random networks where attributes are distributed at random tend to have an attribute assortativity of 0 unless there is a structural component in the network causing for higher or lower levels of nodes being connected rather than a behavioral influence.

After we calculate the attribute assortativity for each day for our network, we also create a permuted version of the network to test if our levels of attribute assortativity is based on individuals' behavior or structural influence of the network itself. To do so, we create 40 permuted networks where the structure remains the same but the proportion of attributes from the actual network that day are randomly reassigned to new nodes. The proportion of each attribute is the exact same from the true network. For this analysis, we look at the attribute assortativity coefficient and confidence interval of the 40 permuted trials to examine the significance of individuals who at some point test positive in Fall 2020 being found

with a shared edge in the network more than if they were randomly found in the network, this visualization can be found in Figure 4 below.
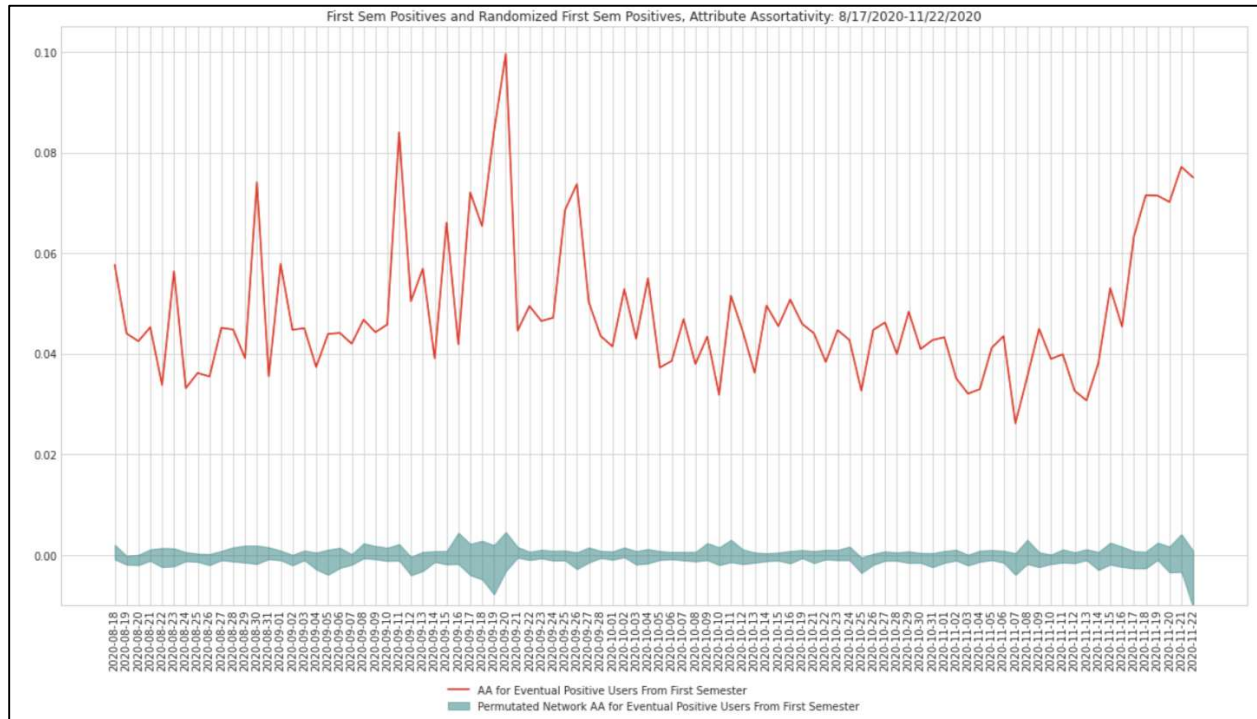

Figure 4 - Attribute Assortativity for First Semester Positives versus 40 Permuted Network Trials

Seeing that the attribute assortativity of the network on every day is significantly higher than the 95% confidence interval of the permuted networks, we then run a new experiment with attribute assortativity by creating a rolling window of positives from the semester. For each day, each individual within the network who tests positive during the Fall 2020 semester and tests positive within 10 days following the interaction is given the attribute of 1. If an individual does not test positive within 10 days following the interaction, the individual is given an attribute of 0, regardless of eventual positivity. We then run the same trial of attribute assortativity of our network each day with these new rolling assignments of attributes in the network. We also again create 40 trials of permuted versions of the network with the same proportions of attributes for that day in the true network to create a confidence interval of a randomly distributed attribute network. This is shown in Figure 5 below.

Though we do not see extremely high levels of attribute assortativity, we would not expect to see within a proximity network of an entire college campus every day to show individuals who test positive during
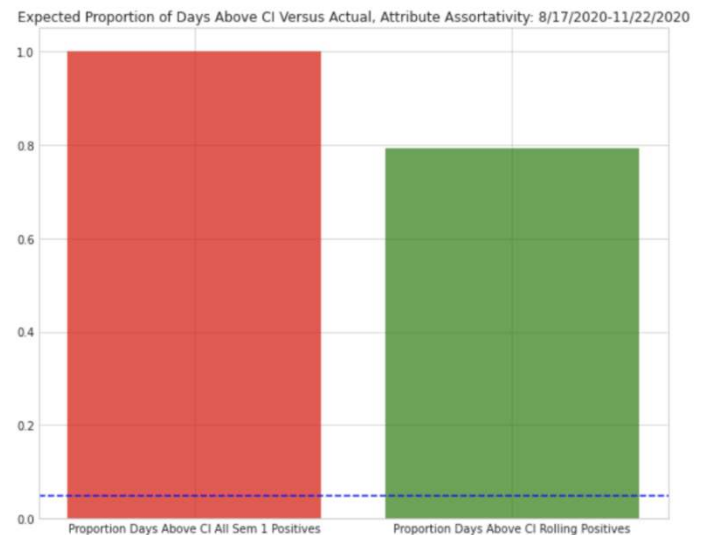

Figure 6 - Expected Proportion of Days Outside of the Permuted Networks 95% Confidence Interval for Attribute Assortativity in Fall 2020

the semester or within a ten day window to only be found on campus with eachother. The key takeaway is that they share an edge significantly more than at random, thus with further analysis close contacts within the network could be identified by analyzing edge relationships of two positive individuals at a more granular scale. The significance of each of the tests is shown in Figure 6 above, where we examine the expected proportion of days to be outside of the 95% confidence interval, which would be 5% versus our true number of days outside of the permuted networks confidence intervals.
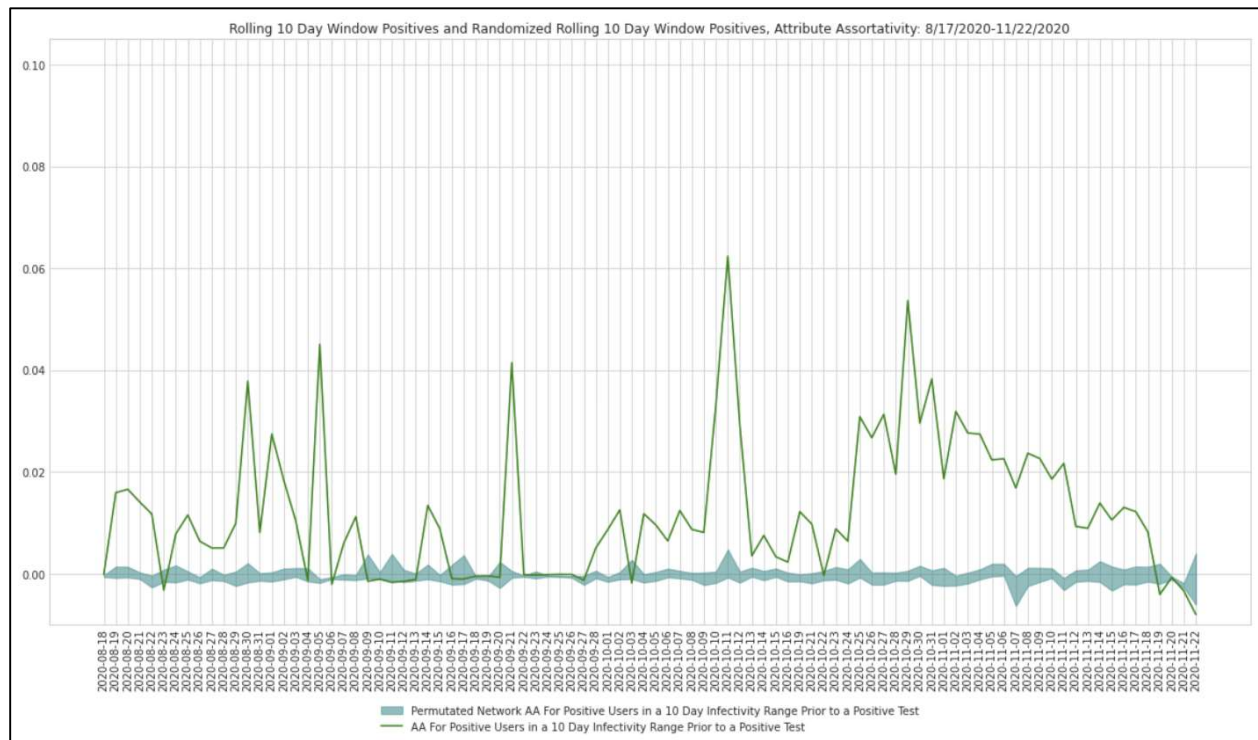
Figure 5 - Attribute Assortativity for the Rolling 10 Day Window Attribute Assigning Scheme in Fall 2020 Versus 95% Confidence Interval for Permuted Network

We can also look at the number of IDs on campus during the Fall 2020 semester to understand how different proportions of individuals can changing width of the Confidence Interval for specific days for each of the above tests. These attribute counting charts can be found in Figure 3 and 4 of the Supplemental Materials Section.

The next step for analysis was analyzing how the trends in attribute assortativity for the 10 day rolling window scheme might be correlated with the case count trends for the Fall 2020 semester. This would test if within a biologically relevant period, were their peaks in individuals listed in the pair-wise proximity network who were known to test positive within 10 days contributing to the peaks in case count directly following that peak. We examine these trends in Figure 7 below. For this chart, the x axis on the left side shows the same attribute assortativity distribution shown in Figure 5 above, overlayed with the case count in Fall 2020 seen in the top half of Figure 1. We also show the smoothed curve of these two distributions using the Savitzky-Golay filter, a commonly used noise-reduction algorithm in the context of biomedical signal processing. The filtering window is 31 and the polynomial order used to fit the samples is 8. These settings were chosen visually by examining the curves produced at a variety of settings.
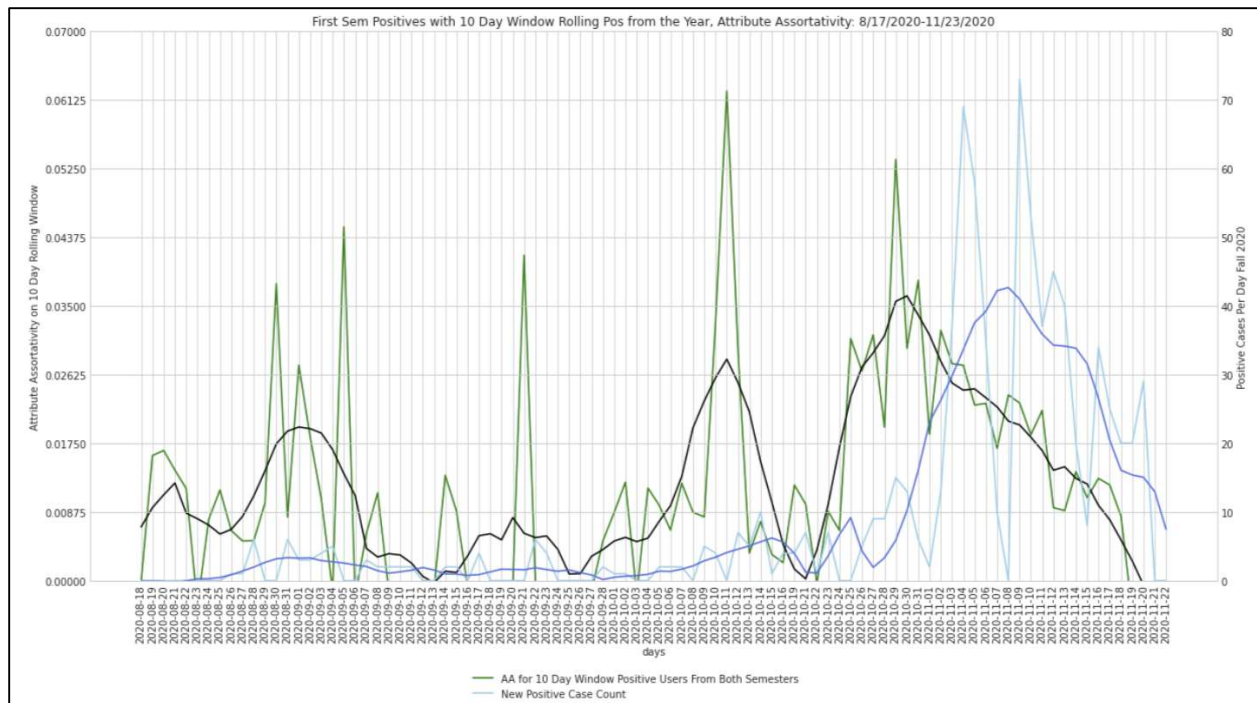
Figure 7 - 10-Day Window Attribute Assortativity with Positive Case Counts Per Day (Smoothed and Raw)

As can be seen in the figure, there is a pattern presenting between the rise in attribute assortativity rates and the short-thereafter rise in case counts. To test this, we use the scipy signal cross correlate tool. This method produces a curve representing a lag time curve versus the correlation coefficient of how correlated both of the signals are with that corresponding lag time. This chart is shown in Figure 8 below. In the chart, the results for the unsmoothed attribute assortativity curve and unsmoothed case count curve is shown in green dots, while the resulting curve for the smoothed attribute assortativity and smoothed case count curve is shown in black. While they shown they generally follow the same distribution, the peak lag time found for the unsmoothed calculations is shown to be 6.57 days and the peak lag time found for the smoothed calculations is 7.58 days.

This lag time represents the number of days the case count curve is following the attribute assortativity curve. These results shown significance using the scipy pearsonr package for Pearson Correlation Coefficient, which measures the linear relationship between two datasets. The result of the Pearson correlation coefficient of the unsmoothed case count per day and unsmoothed attribute assortativity shows a p value of .03354 with the correlation coefficient of .2218. The smoothed curves show a p value of .00018 with a Pearson correlation coefficient of .3807. These results are interesting as it reflects the expected incubation period for the virus, regardless on if the pair-wise interaction data showed true transmission events. Through these results of the proximity network, we see that individuals who test positive and are recorded together in a pairwise interaction are most significantly found together approximately 6 to 7 days prior to an increase in case count. This opens doors to extended analysis of pairwise connections of positive cases, extended analysis of how we could model the potentially risk pairwise interactions, and how that model could be tested to evaluate effectiveness if it were to have been implemented in real time during the Fall 2020 semester.
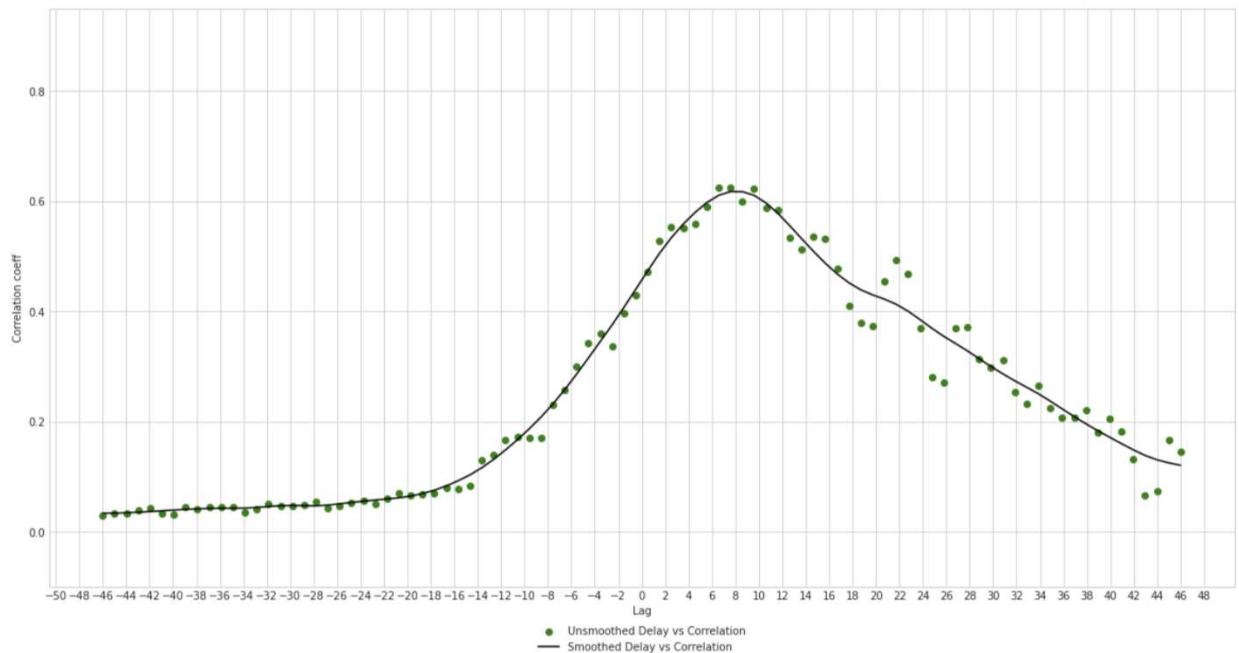
Figure 8 - Lag Time Versus Correlation Coefficient for 10-Day Rolling Window Attribute Assortativity and Case Count

Category 3 Proximity Network Analysis: Using the proximity network as a guide to model the auto-identification of high-confidence potential close contacts

This section will be primarily designed as a discussion of future work. There is still much to analyze with this robust dataset, especially in terms of its potential use as a model building tool. I have plans to create, but have not yet implemented, a model that runs day by day over this dataset. Its main functionality would be to, for each day, account for the neighbors of nodes at one edge distance away (directly connected neighbors). Based on those neighbors for each node, I would like to randomly select from my list of positives for the day from the rolling ten-day window dictionary. I would then like the model to identify high confidence potential close contacts to that selected node that the model predicts will also test positive within 10 days based on this dataset. To reach this point, I need to complete deeper analysis on what the positives within the ten-day periods have in common and find any unique features that define them based on individuals who never test positive or individuals who do not test positive within that ten-day window. Within our dataset, we do have one lineage that was only found on the Colorado Mesa campus (the Peacock cluster), which have 100% sequencing data for from the positive cases tested. This is a cluster I will use explicitly to help construct the model to identify high confidence close contacts.

Once the model is constructed in a beta format, I would then like to run analysis on if we were to have hypothetically had and used this model in real time, how it might've contributed to case counts/response efforts if implemented during the Fall 2020-Spring 2021 school year. This could show as a case study for why a real-time system implemented in Wi-Fi systems could impact close contact tracing efforts.

Conclusion and Acknowledgements:

This paper has been written in the structure of a typical group research project and thus the phrases used to discuss analysis completed has been paired with identifying terms such as "we" and "us". To clarify, the

analysis of this proximity network has been completely individual, with significant mentoring from my Broad Institute team and my UVM supports. I am the only individual who has done any hands on analysis for this dataset and thus all work outlined above has been coded, structured, and implemented by me. That is not to say ideas were all completely independent, I definitely had guidance and advice throughout the process for rabbit holes to go down and others to avoid. For MOCS, the submission was supposed to be mainly related to my work with the permuted network. For POCS, it was more to do with the general network analysis and background analysis shown in the POCS Supplemental Materials section (which was not turned in with the MOCS submission).

Supplemental Materials:
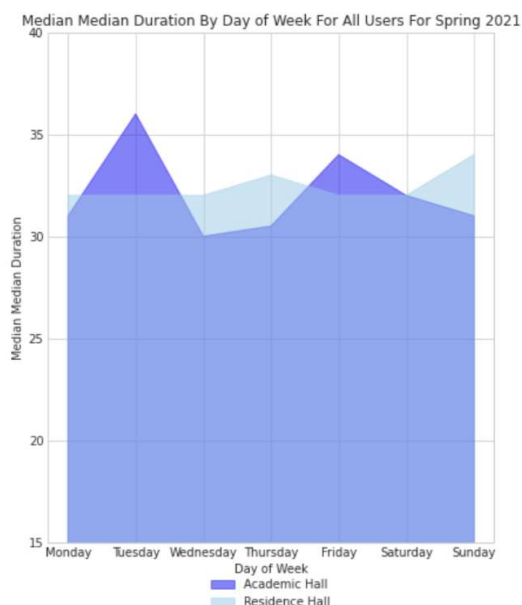
Figure 1 – Median Interaction Duration Median 'ed for Each Day for Spring 2021
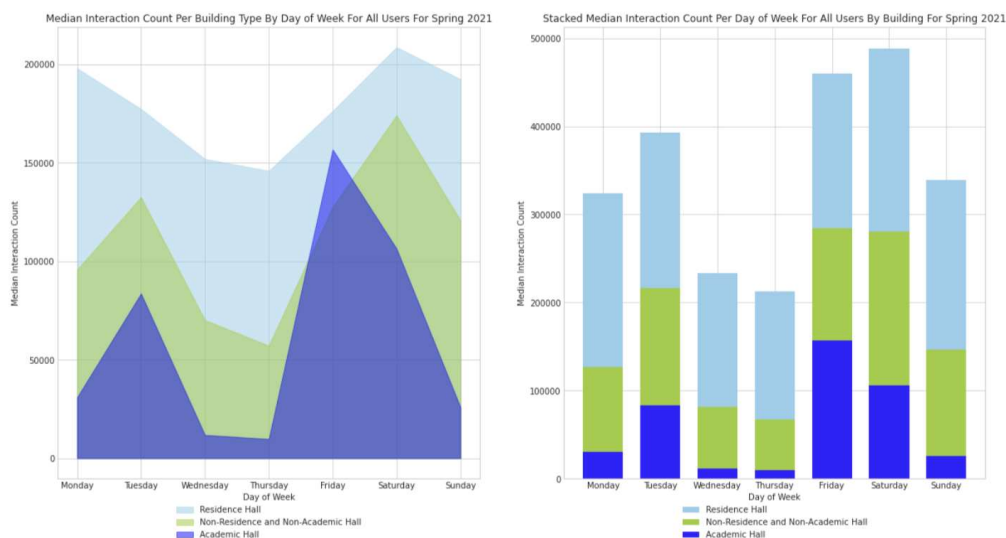
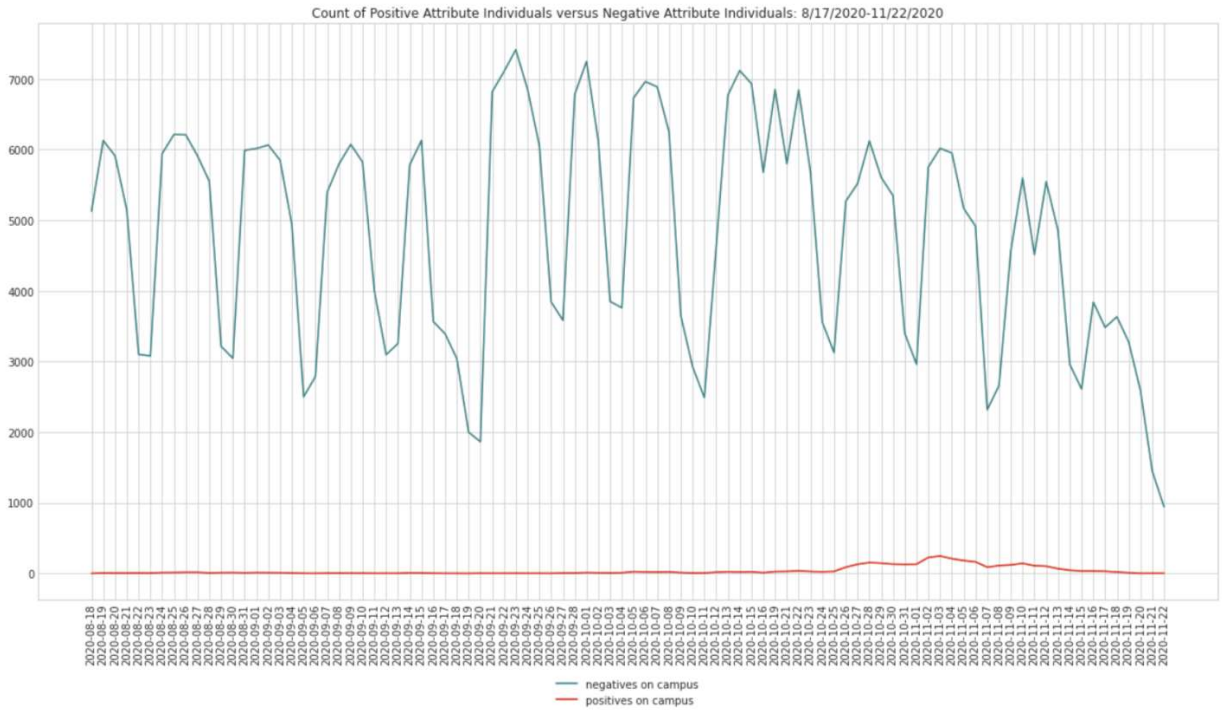*Figure 2 – Median Interaction Count Per Day of Week By Building for Spring 2021*

Figure 3 - Node Attribute Counts Per Day for Fall 2020 in 10 Day Rolling Window Scheme
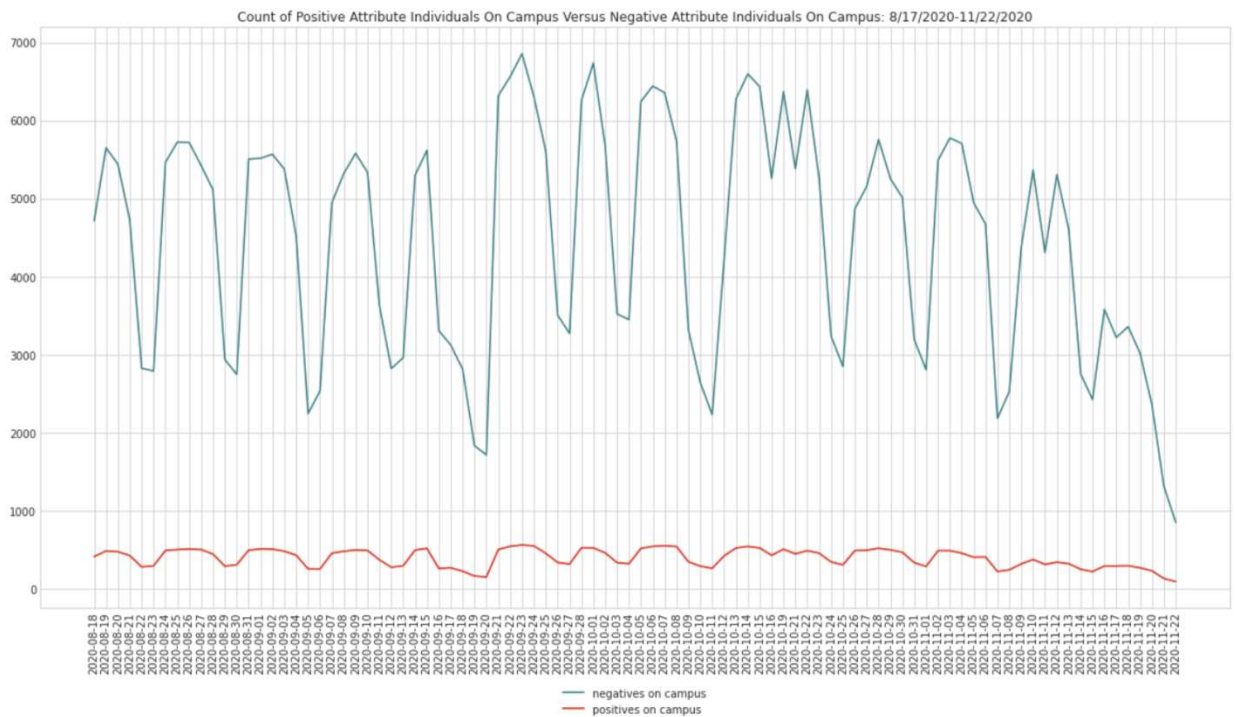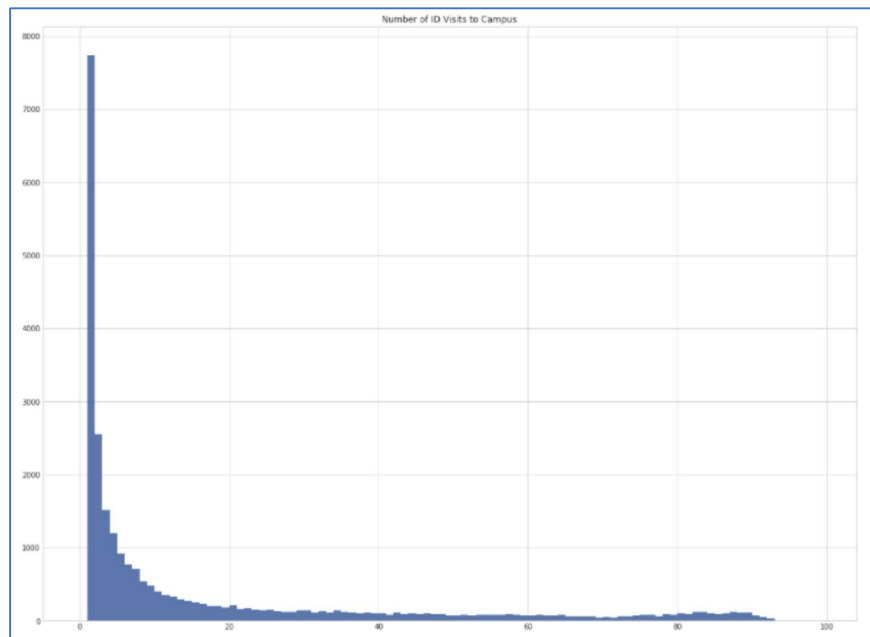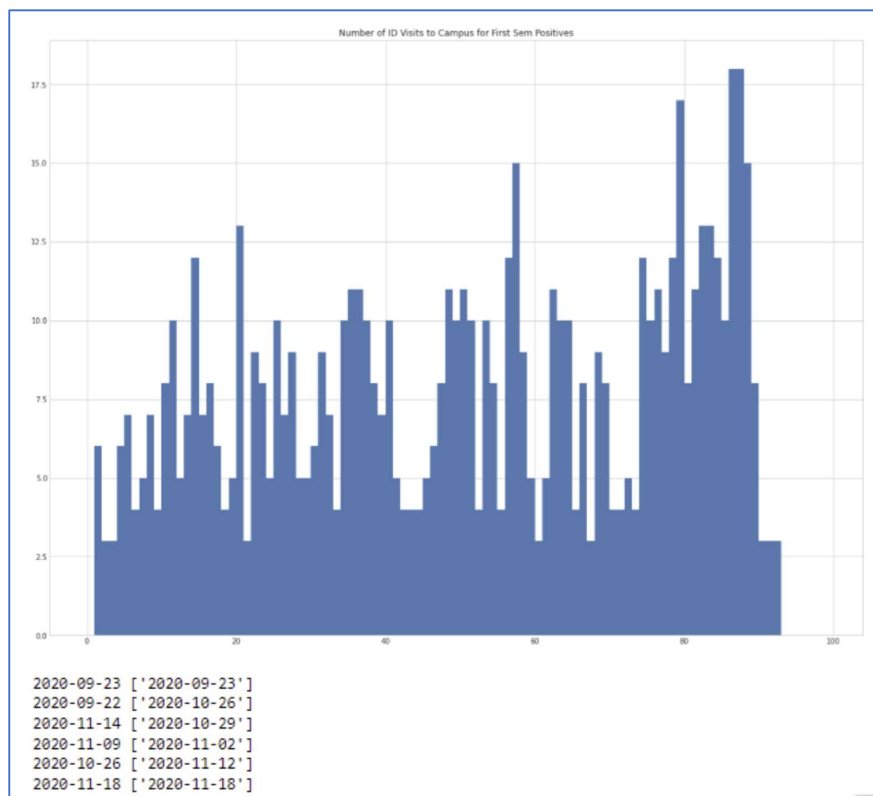
-

Figure 4 - Node Attribute Counts Per Day for Fall 2020 in all Fall 2020 Semester Positive Scheme

POCS Supplemental Materials:

Identifying Guests in the Network for Number of Visits Per ID Per Semester


```
2020-09-23 ['2020-09-23']
2020-09-22 ['2020-10-26']
2020-11-14 ['2020-10-29']
2020-11-09 ['2020-11-02']
2020-10-26 ['2020-11-12']
2020-11-18 ['2020-11-18']
```

Verifying the removal of guests by seeing 2/6 1-time visitors for the Sem 1 positives was on their test date and only one of the 1 time visitors for the Sem 1 positives was in their 10 day rolling window period
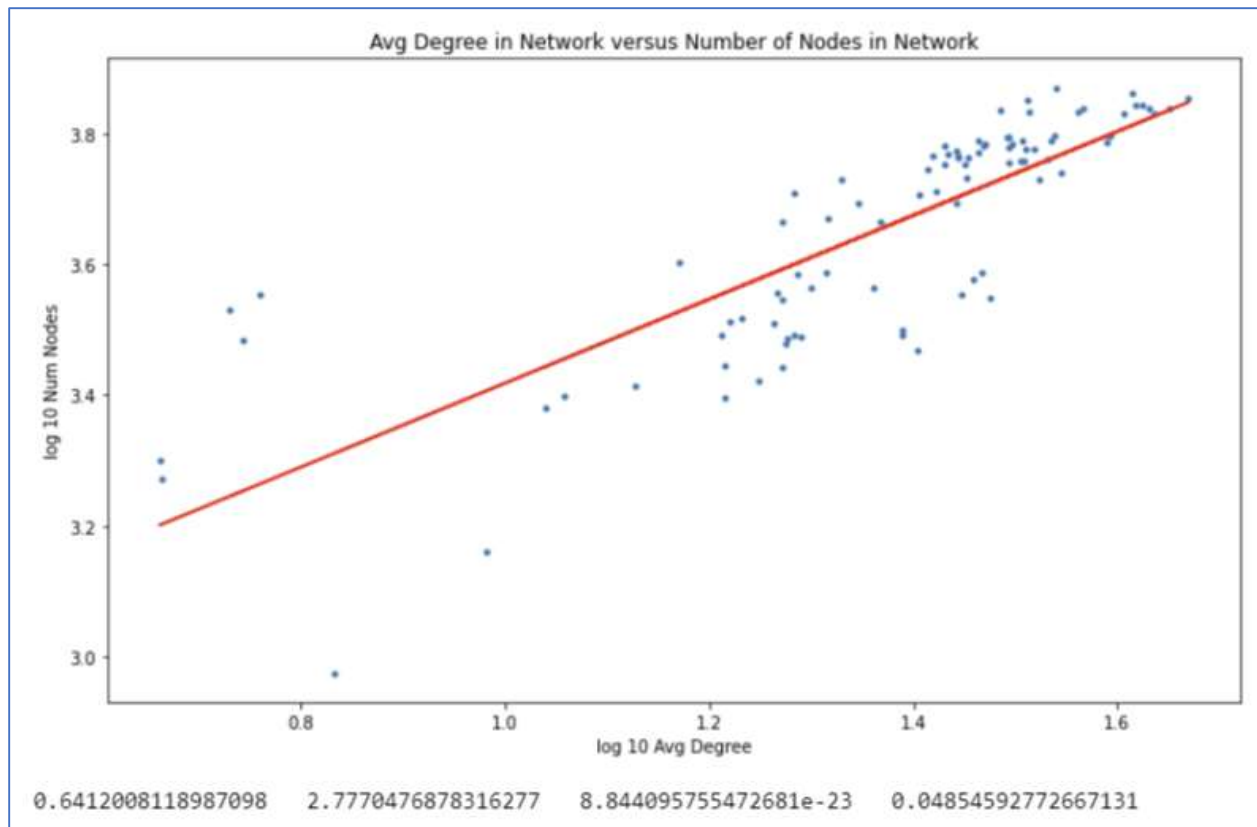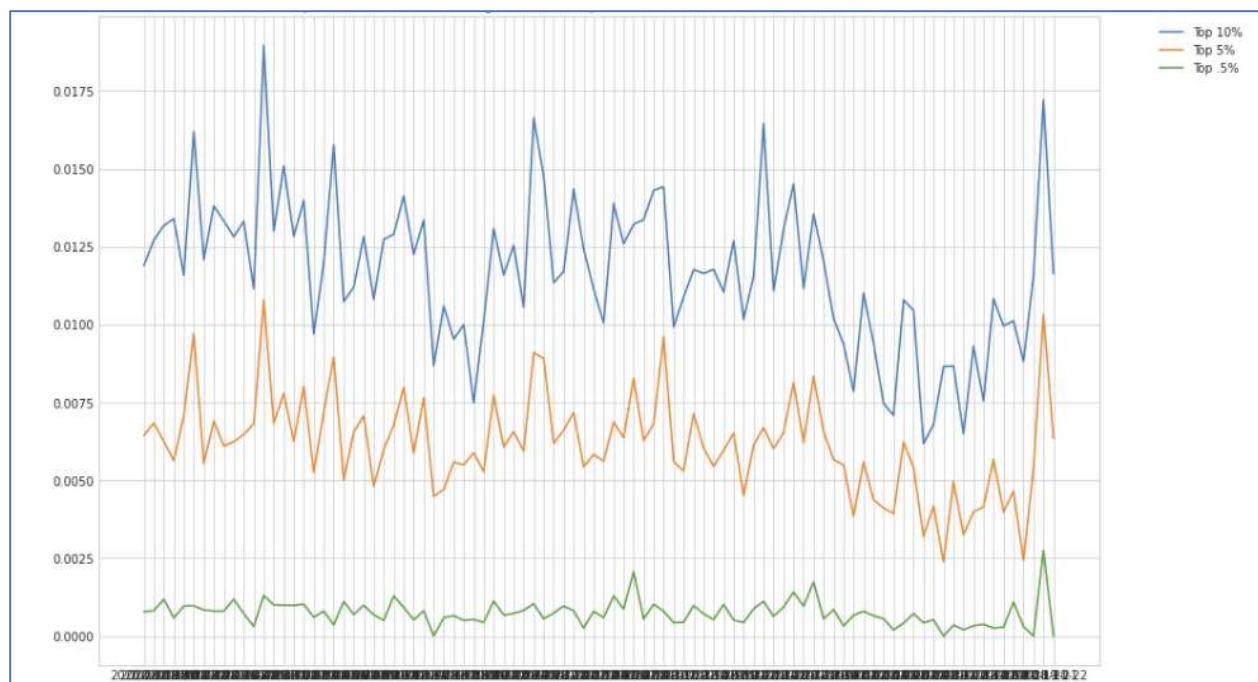
Looking at Log10 Avg Degree versus Log 10 Num Nodes in the Network, each dot is a day in the first semester. Producing Linear Regression line with slope, intercept, p value, and std error
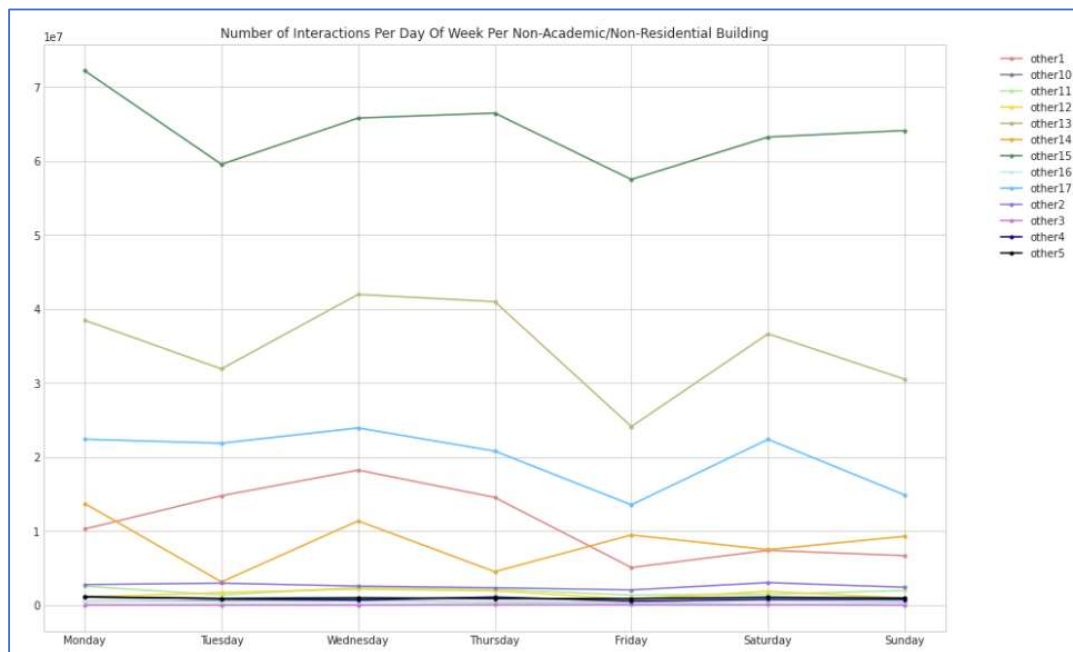
Looking at trend differences in non-academic and non-residential buildings for semester 1
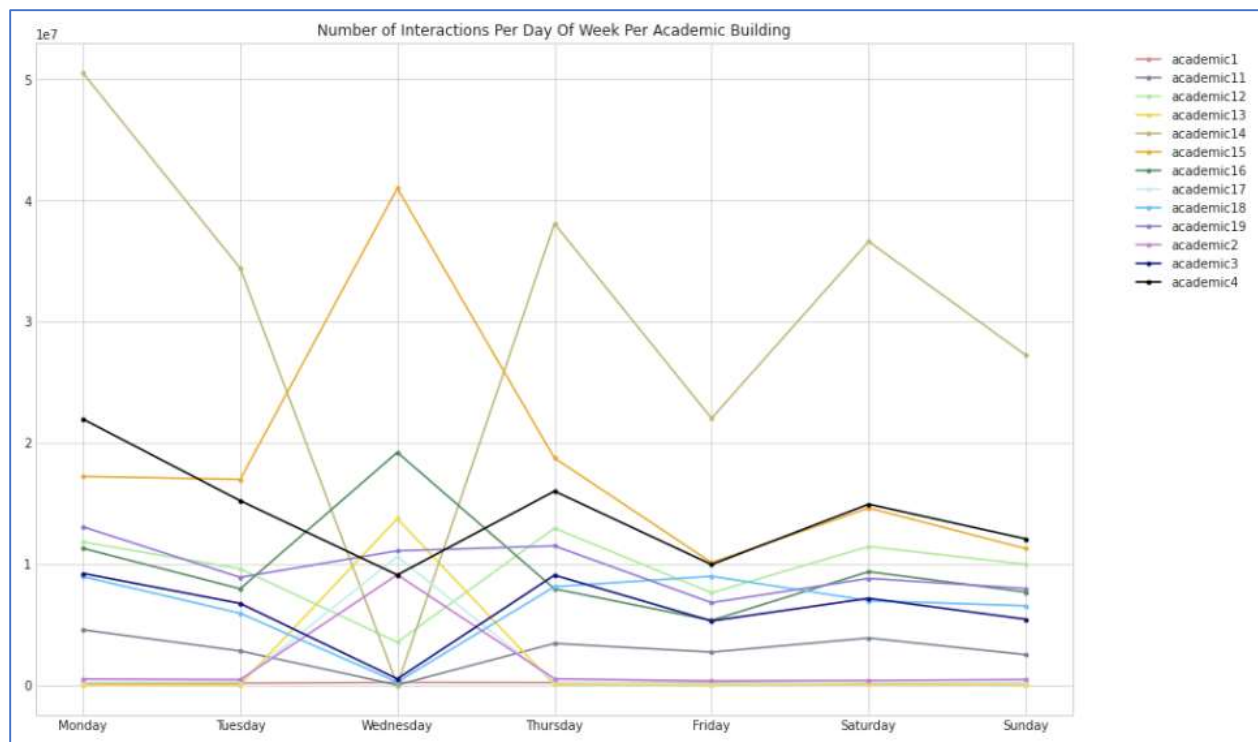
Looking at trend differences in academic buildings for semester 1

Proportion of positive users from first semester in top .005,.05,.1 of users degree distribution for semester 1 per day