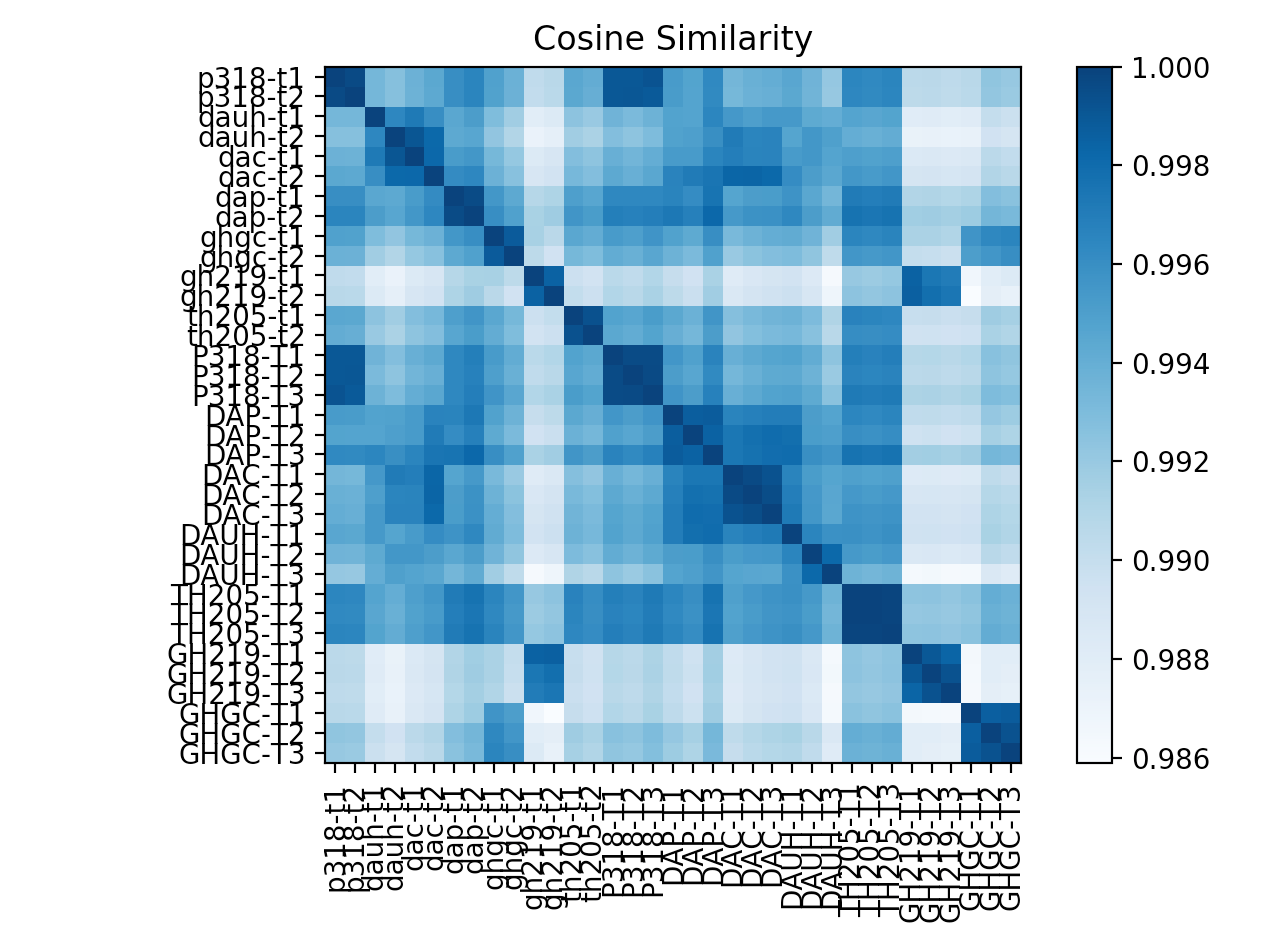


**Computing my own Cosine Similarity:**

In order to interpret this graph, the scale on the right must be interpreted first. The scale shows that the darker the hue/color of the intersection means that the cosine similarity of the two signals has a value closer to one, which means that the intersections are more similar. In the graph, this is shown when the wifi-signals taken from relatively close/similar areas have an outcome that has a darker hue; the intersections that are the darkest, are intersections of the same signals, which is important to have because they represent a somewhat baseline measurement relative to all the other values for comparison. It is also important to note that when capturing data in the upstairs room of Gates Hall, its signals compared to other signals from around campus are very different, which corresponds to our knowledge that Gates Hall Room 219 (gh219) is the most isolated/the farthest place from all of the other places where measurements were taken. Also it should be mentioned that even though, Gimme Coffee (ghgc) is located in Gates Hall, it is relatively similar to all of the other signals than Gates Hall Room 219, since it is an open area down the stairs located closer to all of the other sites where data was taken. In my code to find Cosine Similarity, I just imported the cosine\_similarity extension from sklearn.metrics.pairwise for calculations, however, I also made a function called cosine\_similarity(a,b), where it does the dot product of a and b and divides that by the multiplication of the square root of the sum of the squares of values in a and b. This was done as shown in the instructions.pdf.



**Cosine Similarity of 2 People:**

As shown in this graph there seems to be multiple spots where the data reading seems to be off, however, these are the cases when my signals are compared to my classmate’s signals taken in the same location. It seems that our signals did not match perfectly, but it can be seen that they were quite similar. It seems that our data was relatively similar to each other, and based on the analysis of the graph, it could be said that it would be pretty accurate when classifying the location of my classmate using my data. The results when using Gaussian Native Bayes (Gaussian NB) in order to classify my classmate’s data were this:

|  |  |
| --- | --- |
| Gaussian Native Bayes classification | Classmate’s actual data |
| ['p318-t2']  ['p318-t2']  ['p318-t1']  ['dap-t2']  ['dac-t2']  ['dap-t2']  ['dac-t2']  ['dac-t2']  ['dac-t2']  ['dap-t2']  ['dauh-t2']  ['dauh-t2']  ['dap-t2']  ['dap-t2']  ['dap-t2']  ['gh219-t2']  ['gh219-t2']  ['gh219-t2']  ['ghgc-t1']  ['ghgc-t1']  ['ghgc-t1'] | P318-T1  P318-T2  P318-T3  DAP-T1  DAP-T2  DAP-T3  DAC-T1  DAC-T2  DAC-T3  DAUH-T1  DAUH-T2  DAUH-T3  TH205-T1  TH205-T2  TH205-T3  GH219-T1  GH219-T2  GH219-T3  GHGC-T1  GHGC-T2  GHGC-T3 |

Through this data, it can be seen that most of my data through the Gaussian Native Bayes classification was able to classify most of the data points quite accurately according to the area, however, Thurston Hall (TH205) was never correctly classified. This could be explained through an error in my data capturing because as I tried to get data from inside of Thurston Hall Room 205, both times it was full, and I was unable to go inside for the measurement. I measured the wifi-signal right outside the room, and that is most likely what has caused this error in location allocation. Other than that the classification is very accurate. I think if I was able to capture this data from inside the room, the classifications would have been more accurate. Also if me and my classmate were to do it together at the same area, the results will most definitely have been improved and would be very accurate.