Project 1: Unix and Perl

2. Processing the data (continued)

Task 3. We are going to create matrices that describe the flow of users between the different buildings on campus. Given two buildings, A and B, for each day $i$, let $n_{AB,i}$ represent the number of times people using an access point in building A move to an access point in building B. In terms of our log file, this means someone has associated (or reassocated) with an access point in A and then associates (or reassociates) with an access point in B. Consider only those people who move from A directly to B; that is, we will not count in $n_{AB,i}$ someone who moved from A to C and then to B. Notice also that $n_{BA,i}$, the number of people using B who move to A, will be a different number. To start, you might consider creating a Perl program that will do this for a single day in March and for two specific buildings. Then, repeat the process for all the buildings carrying traffic in March. Right now, we will count the number of transitions, no matter who made them; a single student might be responsible for all the changes between the buildings. If you are feeling ambitious, you might want to instead tabulate $n_{AB,i}^*$, the number of unique people (MAC addresses) that move from A to B.

Next week, we will take these matrices and create maps of the campus based on flow and compare them to the geographic maps of campus. We will also see how the maps change over time. (Recall Lee Wilkinson’s reference to multidimensional scaling; this will be our main map-making tool.)