OEC 2021 PROGRAMMING CHALLENGE PLANNING DOCUMENT

ZBY1 Outbreak – As if 2020 Wasn't Bad Enough!

Team Go

We wanted to break this problem down into different components necessary for the solution to function (and function well). We decided that our two main components would be:

1) **Outbreak Tracking**: To track the overall trend of the virus spread throughout the day
2) **Contact tracing and notifications**: To track and notify the specific people who may have been exposed at the school

Outbreak Tracking Possible Approaches:

The first step was prioritizing our factors for determining who would be more likely to be infected. After reading (and re-reading) the list of criteria that impact the spread of infection, we noticed that some of them overlapped. For example, tracking the fact that the virus can live on surfaces of a classroom can also be accounted for in the fact that students transitioning between classes may be exposed.

We thought about adding our csv data to a database with the intention of writing simple queries to access our data. Due to the size of the dataset, we decided it would be much more efficient to simply store our data in data frames. This allows us to parse our data much more easily and develop much faster given the time constraint.

Our biggest concern was not only tracking the exposure of the ZBY1 virus, but also effectively keeping track of an individual's risk of infection as a result of exposure. With this in mind, we knew we needed to follow set theory, it was just a matter of how to combine our datasets.

We started with the following rough math:

Let A = students, B = teachers, C = TAs, D = infected people

$$P(A \cup B \cup C \cup D) = 1 - P((A \cup B \cup C \cup D)^c)$$
$$= 1 - P(A^c \cap B^c \cap C^c \cap D^c)$$

We later found that a Bayes interpretation got us a much more accurate calculation for probability.

Prior Probability

Likelihood of the evidence 'E' if the Hypothesis 'H' is true

$$P(H|E) = \frac{P(H) * P(E|H)}{P(E)}$$

Posterior Probability of 'H' given the evidence

Priori probability that the evidence itself is true

We later discovered that these calculations were very computationally expensive, since we were iterating though all possible combinations. At this point, we decided to apply a heuristic to our iteration by only focusing on the 5 students most likely to be infected in each class. The logic behind this is that if 5 people are infected, and they can each infect 3 other people, it is possible that most of the class becomes infected for a class of 20 students. We decided to use the 5 highest values rather than covering the entire class to increase code efficiency.

Data Manipulation:
As outlined by our Directors, we were allowed to manipulate our datasets to adhere to any format we wanted, so long as it did not tamper with our analysis. We split our Excel spreadsheet into 4 csv files for each of the pages. We also decided to add an extra column for phone numbers to aid in our notification system.

The rationale behind this is that the school would already have this information in their student and employer records (along with each person's consent), so utilizing each person's contact information for exposure notifications would adhere to any privacy laws or concerns. It would also be the most practical method of notifying. The intention of this notification is to simply report an individual's risk of infection and advise them to avoid further contact with people.

Contact Tracing and Notifications:
Most of the logic for the infection probability is handled in *Outbreak Tracking*. As an additional feature for this system, we thought it would be extremely beneficial to have a method of notifying the individuals who have been exposed during the day. This would report their likelihood of infection on a personal basis and advise them to avoid further contact with others. A notification to an already infected person may be pointless, but there is no way for us to track who exactly the virus has spread to; we can only track the likelihoods. If we can notify anyone who was exposed but may not have been infected yet, then we can break the chain and justify this implementation.