

Simulation of Interactions between Students on Campus

Mingu Jo, Seungjun Lee
Supervisor: Professor David Aldous

ABSTRACT

Each day 25,000 undergraduates and 10,000 graduates often walk around campus and meet different college students. Our basic hypothesis is that interaction between students is based on randomness. The problem of interest is to test such hypothesis, and predict the probability of encounter on campus between different college student on a typical school day. In order to build a simulation model that generate encounters, we collected samples of students' path by conducting an online survey for different college students within UC Berkeley. So far, we have collected approximately 70 paths from each college. Each input data representing individual path will be compared to data from other colleges, thereby finding interaction points with given inputs. The simulation calculates the overall probability of interaction between students from different colleges, and the location with highest probability. The result showed that students in certain colleges were more likely to meet other college students at random. Also, we noticed that the location with highest frequency of interaction did not reflect the students' majors.

Introduction

People on a university campus, especially students, tend to follow certain routes on campus. Such routes mainly depend on their class schedules. However, department buildings and classrooms are concentrated at particular locations rather than located randomly. Therefore, given the assumption that students travel on a campus according to their schedules, it seems plausible that students in certain colleges, such as Haas school of Business, have lower chance of meeting students from other colleges. To demonstrate such idea, this study is designed to see the chances of students meeting other college students at random around campus based on their individual path on campus.

This research will provide interesting information regarding the chance of meeting different college undergraduates by running a simulation model based on real world data.

OVERVIEW OF ONLINE SURVEY

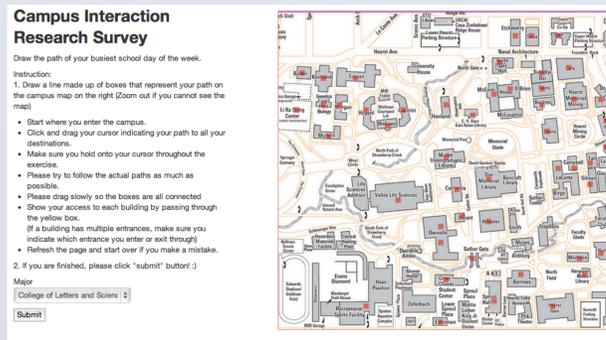


Figure 1.Link : <http://ec2-184-169-194-161.us-west-1.compute.amazonaws.com/>

Data Collection and Simulation Process

Methodology

Our research methodology requires gathering individual paths that students use according to their colleges (or majors) in order to visualize their interactions through simulation. Bootstrap process shall be utilized for this research project to generalize the results from each simulation.

Data Collection

Data set will be collected through online surveys targeted to UC Berkeley undergraduate students. The relationship between number of observations for each college will represent the major population proportion in the real world. Initially, the campus map will be represented by a matrix with 60 rows and 80 columns (i.e. matrix size = 4,800). We classify the majors into six groups: College of L&S, College of Chemistry, College of Engineering, College of Environmental Design, Haas school of Business, and Others. Consequently, students will specify their college (or major) and input their usual path on campus in regular sequence. The paths will be recorded in separate vectors.

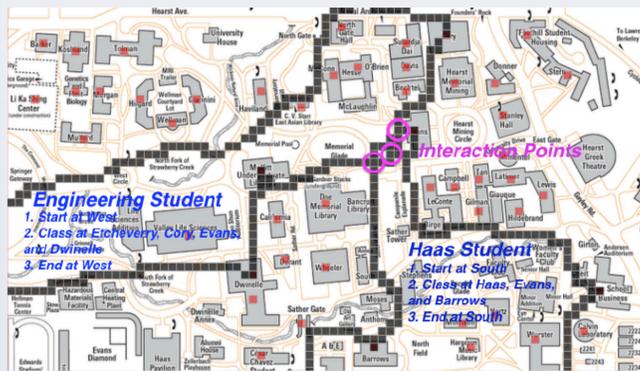


Figure 2. Simulation of interaction between an engineering student and haas student.

Simulation Process

For each simulation, we will calculate the interaction between six students, each representing different colleges. We shall utilize probability sampling to include 1 participant from each group, selected randomly from within the groups, in order to have a good representation of all colleges. For the purpose of simplification, we will regard "time" as binary variable—AM and PM. Accordingly, each of the six vectors representing individual path will be divided into AM (first half) and PM (second half).

In this research, we utilize *Bayes' rule* in order to estimate the probability of interaction between students in different colleges. The probability of interaction between student A and student B will be calculated in the following method:

$$P(\text{interact}) = P(\text{interact} | \text{AM}) + P(\text{interact} | \text{PM})$$

where

$$P(\text{interact} | \text{AM}) = \frac{(\# \text{ of matching AM points})^2}{(\# \text{ of AM points for A} * \# \text{ of AM points for B})} \text{ and similar for } P(\text{interact} | \text{PM})$$

Results

In each simulation process, we calculate the probability that two students interact within the simulation. For simplicity, we define *interaction* between two students as "using a path at the same time frame, either AM or PM." Under this definition, we collect the probabilities of interaction calculated using different paths in each simulation.

We use bootstrap function in order to repeat simulation 10,000 times. Subsequently, the final result will consist of the mean and standard deviation of collected probabilities along with the location with highest probability of interaction.

The result will be printed in the following way:

List 1 (L&S and Engineering) ← Compare the two groups separately

- Mean of probability
[1] 0.03707534
- Standard deviation
[1] 0.07320441
- Location with highest frequency
[1] "1jGnAgwSxYYA"

List 2 (Engineering and Haas)

- Mean of probability
[1] 0.02667699
- Standard deviation
[1] 0.0641065
- Location with highest frequency
[1] "bJQLwphlWHSK" "zMpP1IRJMSfl"

List 3 (L&S and Chemistry)

...

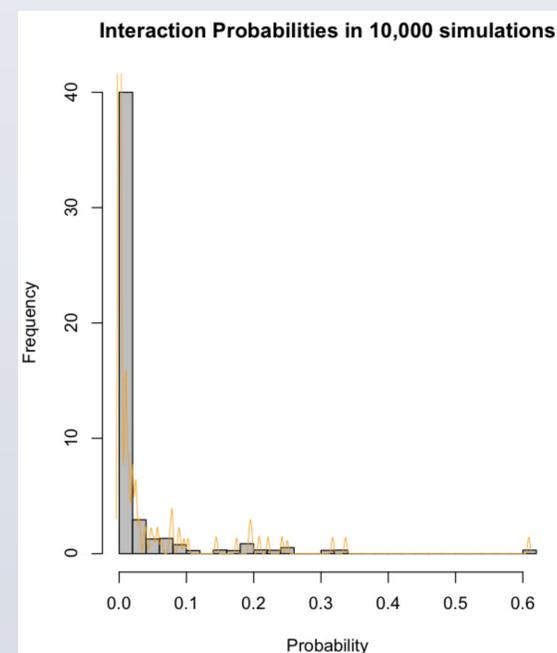


Figure 3. Histogram of interactions between L&S students and Engineering students

Conclusion

We investigated the relationship between students' interaction and their majors by running a simulation model multiple times, each independent to one another.

Following are the current conclusion and limitation of our project:

1. Students in certain college were more likely to interact with students in other colleges.

- Probability of interaction was generally higher for L&S students.
- College of Engineering students and Haas students were less likely to interact with each other.
- Location with highest frequency of interaction did not reflect their majors—the interactions occurred at rather random locations.

2. Overall probability tend to increase as we increase the size of dataset.

- Our current result is based on only 70 observations, thereby decreasing the frequency of interactions.
- We shall acquire at least 300 observations, retaining the population proportion across colleges in order for the sample to represent the population well.

Future Research Considerations

- We shall analyze each path as time series data in order to extract meaningful statistics that may be influenced by time variable.
- We should provide better user experience on the online survey for collection of more data.

Contact

For questions and inquiries about our research, please email us at mingu08@berkeley.edu or lsj0818@berkeley.edu.

Mailing Address:
2335 Dwight Way APT 202, Berkeley, CA 94704

Thank you.

