Demographic Characteristics of Students Who Do or Do Not Post in an

Undergraduate Engineering Online Discussion Forum

Authors

Affiliations

1. Introduction

The current schooling norm is that students can only get help during the class hours and limited office hours. They are not in touch with their instructors or classmates outside of the class. The introduction of an online discussion forum for a class helps students stay in touch with their peers, teaching assistants (TAs) and even instructors. Such a discussion forum also provides a public space where ideas can be exchanged, questions can be shared, and information can be stored. Instructors and TAs are able to help students clarify doubts with fewer time and space limits. These forums can then be considered to be a social network, where Social Network Analysis (SNA) and statistical computing techniques can be applied for studying the interaction of the students and levels of participations.

Here we propose to apply SNA and statistical computation techniques to an online discussion forum to discover how frequent a student participated, and to discover how demographic characteristics related to student participation. This statistical analysis will help instructors better understand relations between students’ participations and their backgrounds.

1. Research Questions

The following two research questions guide this work:

* 1. How do students’ demographic characteristics relate to discussion forum participation?
  2. Among students who participate, how do students’ demographic characteristics relate to the number of their posts?

Among the demographic characteristics I am focusing on, we expect female students are more likely participate in the discussion online forum compared with male students. We also expect that international students are quieter when discussing online. Since the course material is closest to mechanical engineering students, we expect more ME students get involved in online discussion than other major students. In terms of participants, we expect that the density of female students of higher number of posts should be over male students.

1. Related Work
2. What are Online Discussion Forums?

Online discussion forums can be accessed via internet by participants in a course[5]. It allows individuals to hold conversations with each other, teach assistants and instructors by posting text-based messages. Everyone can post and check it from any place at any time, so it is a very useful source across traditional and online platforms. This study focuses on a sophomore-level dynamics and vibrations class that incorporated active, blended, and collaborative learning strategies. The corresponding course discussion forum can be accessed by any students and instructors. Class students communicate with their peers of course material, such as homework problem solving.

1. Social Network Analysis (SNA)

Social Network Analysis (SNA) is the use of network theory to analyze social networks. This turns the network to a visualized social network diagram, where nodes represent the individuals and edges depicting the relationship between individuals. In this study nodes are represented as circles and edges as lines. Statistical computation is aiming at the design of algorithm for implementing statistical methods. It is the application of software open-source to statistics nowadays. In this study, we especially use bar graph to compare results from different variables.

Previous studies showed that SNA is a common tool to evaluate students’ online discussion performance. He[2] uses SNA software to visualize students’ online discussion participation networks. That paper evaluates individual student’s online discussion performance based on SNA pagerank and in-degree out-degree centralities. However, based on how actively students participate in online discussion, they focus on analysis of academic performances instead of demographic characteristics.

There are also some previous studies about demographics of social network users. For example, in a case study, Yeh etc.[3] demonstrates demographics of users in terms of gender, nationality and ages. He announced that more than half users are female (53%) and at age of 16-30. Ke etc.[1] completed an online learning across ethnicity on learning interaction participation. They concluded that students’ minority status was associated with lower satisfaction with the web-based distance learning. Our study’s sample is from a specific online discussion forum instead of a famous international online courses website. Andrew[4] built a conceptual framework for demographic groups resistant to online community interaction. From the previous study as you can see, the study of demographic characteristics of students who participate in an online discussion forum is very limited.

1. Research Method

In this section, we depict the sample data and the features we used for analysis.

* 1. Data Collection

This mechanical engineering(ME) course forum assigns each post a unique id. As long as students write a message online, their posts are recorded under the student’s id. Then, part of the post related information is extracted into a csv file by a software tool. This includes the person id, which person this post is communicated with, and post time. In on our discussion forum situation, the instructor always starts a new thread by posting homework problems or lecture notes, and students post their doubts or opinions under instructor’s posts. Since we are only interested in students’ online social behavior, the posts started with an instructor has been removed.

* 1. SNA

There are several SNA methods that one can consider when analyzing a social network. Particularly in this paper for the context of evaluating student’s online discussion performances, we will use out-degree centrality to count the number of messages sent by a student. If a student acquires high numbers in out-degree, it indicates that he or she is more active in sending messages to others. There were 954 unique students participated in our course’s online discussion board and total number of messages was 1861 in three academic semesters from 2015-2016. These included students’ posts and replies, but initial posts from instructors and teaching assistants were not counted.

* 1. User Profiles

We collected all unique user profiles from surveys. We retrieved the following user profiles data:

* + - * User Personal data: Gender, Nationality, Ethnicity, Major, and ADCI(Advanced demographic C)

Gender is divided by Female and Male. Due to this study happened in the USA, nationality is divided by international students and Americans(citizens). By ethnicity, Americans are categorized by White American, Asian American, Hispanic or Latino, African American, Unknown, and 2 or more races. In this case, we did not dig into international students’ ethnicity, as international students only occupy a very limit amount of total. As mentioned previously in online discussion forum session, this class is about dynamics and vibration, which is required for mechanical engineering(ME) students, selective for Agriculture Engineering(AAE), Nuclear Engineering(NE). Students from Multidiscipline Engineering Study, Biomedical Engineering can also take this class for credits, so I summarize them as one Other category.

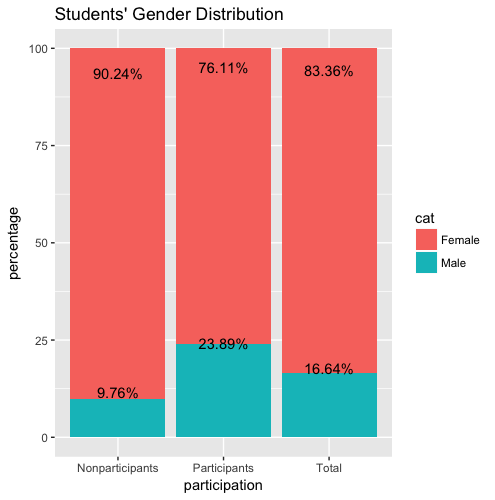
Models based on demographic attributes are important since it helps to determine the connectivity based on social attributes[7]. We got all user personal data based on students’ answers to a survey. It is imported as a raw data. Then, a filter is applied based on consented status. We totally got 739 consented students, among which 360 are participants, and 379 are nonparticipants. Three data frames of user personal data plus out-degrees are generated via R. One is of consented participants, one is of non-participants, and one is of total students. Each data frame contains belonging students’ demographic information. Manual validation check is applied. For example, we manually counted the number of students who post below 5 times, and it matches the data R calculated. Statistical graphics are generated for each perspective of selected demographic characteristics. Because we are also interested in finding if there is a demographics pattern among how frequently a person posts online, later we focus on analyzing participants. Density plot of gender versus number of posts is conducted.

Last but not at least, Chi-Square goodness of fit test is applied. This test allows us to test whether the observed proportions for a categorical variable differ from hypothesized proportions[2].

1. Results and Analysis

First, we give an overview of sample data we collected, including gender distribution, ethnicity distribution, major distribution, and ADCI distribution. We compare the result with “Chi-Square goodness of fit” test. Then we looked at the gender and students’ performance distribution among the number of posts.

1. Participations vs. Nonparticipations
   1. Gender Distribution



*figure 1 Student gender distribution among participants, nonparticipants and total.*

From the distribution graph above, we can see that 23.89% of discussion participants are female students. The percentage value is less than that of male students due to the unbalanced ratio of this engineering class. From the third column, we can see that the whole class only have 16.64% of female students. More than half of the female students get involved in this online discussion forum more or less. There are 76.1% of participants are men but 83.36% of total class students are male students. Therefore, female students are relatively more active in participating in online discussion forum. The gender distribution is similar to the result of many previous studies[3][6], where they declared that women are more active than men on social network media and online discussion blog. It also matches the Donovan’s observation[8].

* 1. Ethnicity Distribution

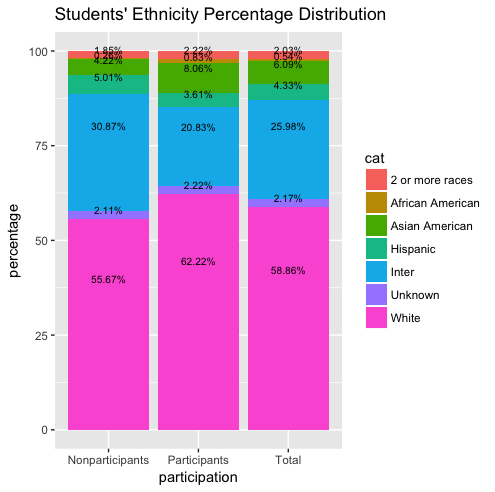


figure 2.a

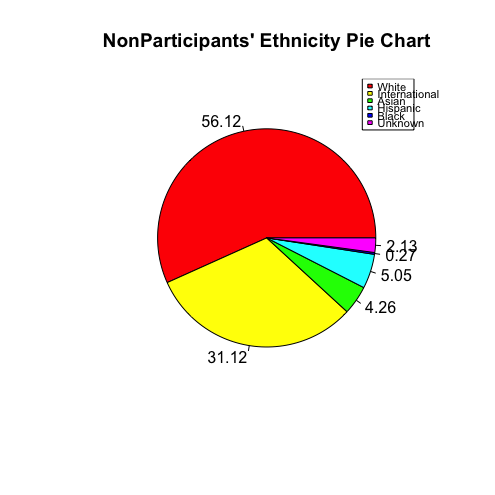
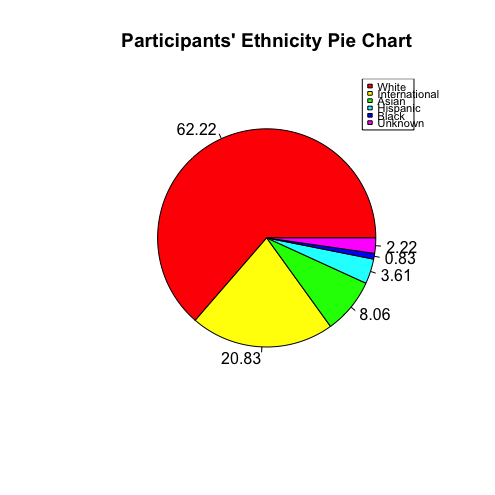


figure 2.b figure 2.c

*figure 2 Students’ Ethnicity Distribution. Figure 2.a, the distribution is of stacked histogram. Figure 2.b, 2.c are pie charts of ethnicity and nationality percentage distribution of participants and nonparticipants.*

Figure. 2 shows the nationality and ethnicity distribution of the sample data. Our total data shows that the white American people dominate the population, followed by International students, Hispanic, and Asian Americans. There are 58.86% of the class students are white American. Among all participants, 62.22% are the whites, and 55.67% of nonparticipants are white people. From these data, we can see that white Americans over-represented. Next, from column three, we can see that 25.98% of the class students are internationals, but this ratio decreases to 20% in participants column. As a result, international students look under-represented in this forum. Moreover, there are 4% of the class students are Hispanic or Latino and 6% are Asian Americans. The ratio of Hispanic students in participants column decreases to 3.6% but the ratio of Asian Americans increases to 8%. Then we can say that Hispanics less involved but Asian Americans are more involved in this online forum. Therefore, White Americans and Asian Americans are over-represented but international students and Hispanics are under-represented. This matches one of the previous work[3], where they declared that minority status has lower satisfaction with the web-based distance study.

* 1. Major Distribution

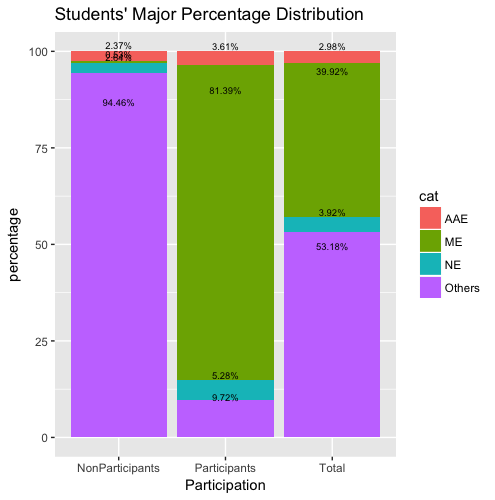


figure 3.a

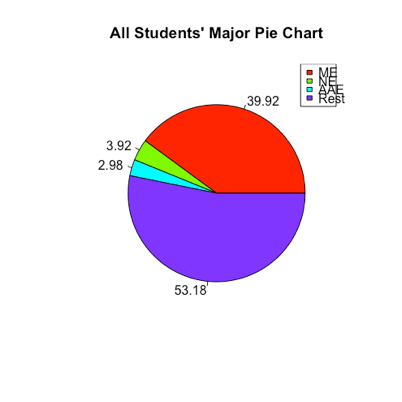
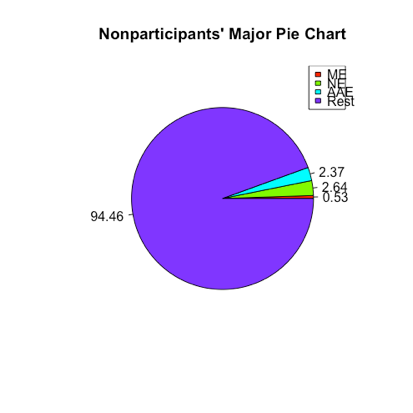
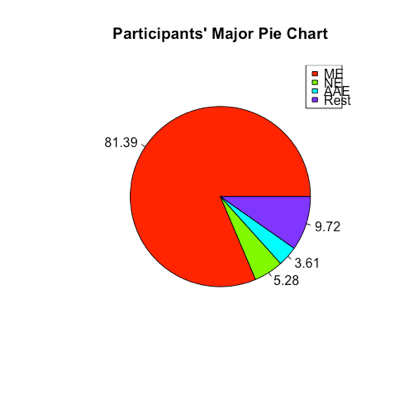


figure 3.b figure 3.c figure 3.d

*figure 3. Major Distribution among participants, nonparticipants and all students. ME represents Mechanical Engineering, AAE represents Agriculture Engineering, NE represents Nuclear Engineering.*

*Figure 3.a is the stacked histogram of major distribution. Figure 3.b, 3.c, 3.d are pie chars of major distribution for participants, nonparticipants and all students respectively.*

Figure 3 shows the major structure of the sample data. This online discussion forum is designed for a dynamic course, so all students are directly from mechanical engineering(ME), Agriculture Engineering(AAE), Nuclear Engineering, and other engineering departments, where students can take this course as a selective. The statistical computation shows that ME are the most involved than other majors, where only 2 out of 237 students didn’t post online. Beyond that, students who take the course as a selective don’t prefer to participate in online discussion at all.

1. Among Participants:

1. Gender Distribution

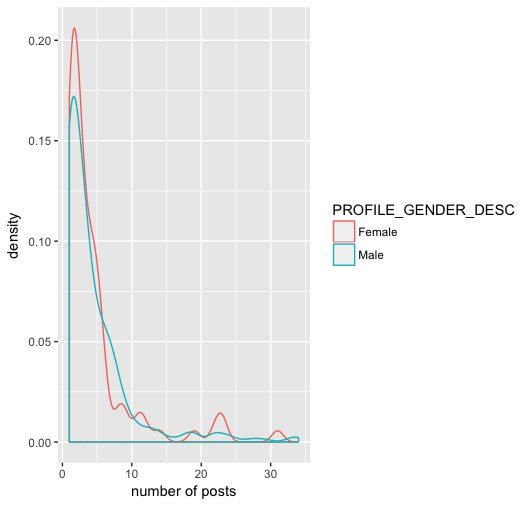


figure 4 Gender Distribution among the number of posts

The figure above shows the gender density trend among all participants. We can obviously see that women have a higher peak value at lower number of posts (0 - 5) and medium-high posts (20 - 25). Here we didn’t count one student, who posts 150 times over one semester, in order to keep the trend line more appropriate to describe the situation.

2. Grade Distribution

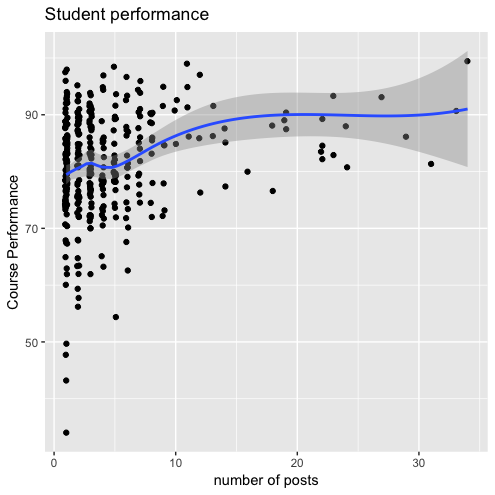


figure 5. Student Performance distribution based on scores.

Figure 5 describes the relationship between the number of posts versus students’ performance scores. The score is a 100 base, and higher score means that students did a better job in the class. We briefly could see a better academic behavior associated with the more active a student is involved.

1. Discussion

From the results we got above, we are able to conclude that gender, ethnicity and majors relate to discussion forum participation. Moreover, among students who participate, women and students who have a strong background of the course material feel more comfortable in posting on online discussion forum.

Due to the fact that we only design our research based on consented sample data, it doesn’t contain all students enrolling the class. However, consented students take up 78% of all students, so we claimed our sample is valid and persuasive.

1. Conclusion

In this paper, social network analysis combined with statistical graphics and validation check have been used to understand demographics of students who do or do not participate in online discussion forums. It is shown that female students are more likely to be involved than their male counterparts. Also, White and American Asians are overrepresented but international students are underrepresented in the engaged group.

1. Future Work

This work extends our knowledge of who uses online collaboration tools, and future work will analyze the content of posts and explore the influence of forum participations on grades via a regression model.

1. Reference
   1. P.Suraj and V. S. K. Roshni, “Social Network Analysis in Student Online Discussion Forums”, *2015 IEEE Recent Advances in Intelligent Computational System(RAICS)*, 2015.
   2. P. He, “Evaluating Students Online Discussion Performance by Using Social Network Analysis”, 2012.
   3. C. Yeh, Y. Lin, T. Lin, “Demographics of Social Network Users- a Case Study on Plurk”, 2012.
   4. D. Andrews, J. Preece, M. Turoff, “A Conceptual Framework for Demographic Groups Resistant to Online Community Interaction”, 2001.
   5. Engaging students for success in calculus with online learning forums
   6. “Analysis on the General profile of users on Plurk.com”
   7. “Demographic and Structural Characteristics to Rationalize Link Formation in Online Social Networks”
   8. “Characterizing user behavior and information propagation on a social multimedia network”

[1] F. Ke and D. Kwak, “Online learning across ethnicity and age: A study on learning interaction participation, perception, and learning satisfaction,” *Comput. Educ.*, vol. 61, no. Supplement C, pp. 43–51, Feb. 2013.

[2] “What statistical analysis should I use? Statistical analyses using SPSS,” *IDRE Stats*. .