

To Heed Or Not To Heed: r/UMD's Effect On Decision Making for Students

Aniruddh Mutnuru
amutnuru@terpmail.umd.edu

Shashank Singh
sps2026@terpmail.umd.edu

Tanay Shah
tanay11@terpmail.umd.edu

Abhinav Mukherjee
amukher1@terpmail.umd.edu

1 Introduction

In today's digitally interconnected world, students are increasingly seeking guidance and insights regarding various aspects of their academic journey, including course selections, housing options, and other related decisions from online platforms. One specific platform is Reddit which has become a consistently prominent forum where students exchange information, share experiences, and seek advice from peers. Our research project seeks to explore the significance of Reddit as a source of information and its influence on students' decision-making processes, particularly regarding academic courses and future housing choices.

The central problem we aim to address in our project revolves around the need to centralize diverse information relating to students' needs, such as class recommendations, housing options, and other relevant considerations. While there are administrative resources for students, the accessibility and relatability of information provided by fellow students on platforms like Reddit offer a more compelling and relatable perspective that informs other students on how to make these decisions. Therefore, understanding the actual significance of online peer input in this decision-making process is crucial.

Our project seeks to address two primary questions:

1. How does information obtained from Reddit comments, posts, and upvotes influence a student's decision-making process?
2. What is the impact of Reddit discussions on students' academic and housing decisions?

Highly upvoted comments and popular posts on Reddit serve as indicators of widespread agreeability and relevance, potentially leading to students' perceptions and preferences. By analyzing various types of content shared on Reddit, including personal experiences and recommendations shared via explicit comments and posts, and implicit upvotes and shares, our research aims to uncover the mechanisms through which Reddit influences student decision-making in the digital age.

In conclusion, leveraging Reddit communities like r/UMD as sources of information can significantly impact students' decision-making processes regarding academic and housing choices. By understanding the dynamics of online platforms like Reddit and their influence on student behavior, educational institutions and academic advisors can better support students in making informed decisions and fostering a sense of community and collaboration among peers.

2 Literature Review

In the ever-evolving digital landscape, platforms like Reddit have emerged as significant sources of peer advice, especially influencing student decision-making in academic and housing matters. This initial literature review explores the burgeoning role of Reddit, focusing on how it shapes the academic journey and living decisions of students, particularly within university communities such as r/UMD.

The first is [1] ["Identifying the social signals that drive online discussions: A case study of Reddit communities"](#) from the Rensselaer Polytechnic Institute published in 2017. This group of researchers developed models that accurately predict which Reddit comments and posts will be the most popular, and provided an analysis of why certain posts and comments gain the most popularity based on the subreddit they are located within. This research is similar to ours and we would incorporate similar tactics in order to derive our own conclusions. For example, they relied on sentiment, relevance, and content analysis specifically made for Reddit by looking at subreddit rules, characteristics, comments, upvotes, and users' karma. Our research would include the same types of analysis on the same types of features in order to assign an influenceability score. This research provides us with the different methodologies of conducting our research to completion should we choose to do so. Additionally, we could confirm the paper's research results and reasoning about users' influence on the discussion and its popularity outcome. This furthering of their research ties directly into our research question and is good inspiration to rely on should we need it.

The next article we reviewed is [2] “[Predicting Reddit Post Popularity Via Initial Commentary](#)” by Andrei Terentiev and Alanna Tempest. In their research, they used various machine learning model techniques like K-Neighbors, Perceptron, SVM, and Logistic Regression which turned out to be the most accurate. Essentially, these models were trained on the following features of posts with at least 10 comments: comment length, sentiment analysis, and “polarity” which they define as the maximum sentiment analysis score minus the minimum sentiment analysis score. These tactics are ones we can employ in our research to determine influenceability. This article differs from the previous one because of its specificity by narrowing down the features used to determine the resulting sentiment and popularity. Both are strikingly similar to our own research idea signaling its feasibility.

The third article in our literature review taken into consideration is [3] “[Using Machine Learning to Predict the Popularity of Reddit Comments](#)” from the United States Military Academy. In addition to using similar tactics as the previous two articles, they also focused on a more concrete definition of popularity (top 25% of comments within each subreddit), and they utilized the subreddit `r/datasets` in order to gain the dataset for their research. This facet of their research provides us with information on collecting our own data should we continue our research for we could inquire about how data is captured from subreddits so we may employ those methods on `r/UMD`. They also employed pre-processing techniques like modifying comments so they all start with a lowercase letter in order to capture an accurate sentiment per comment. Their accuracy for the models was between 42.0–52.2% which leaves room for improvement in our research. This study is unique because it focuses solely on comments rather than comments and posts or another combination with posts which introduces new questions that can be researched further, and it provides us with an outline on how to conduct our research.

3 Approach

In order to confront our problem after reading previous works, we first needed to curate our own dataset of both the subreddit’s features and the user-behaviors and opinions. To do so, we collected two main datasets: one from `r/UMD` using the `PRAW_API`, and one from our own collection via a survey of many UMD students.

3.1 PRAW_API Data

We aimed to extract data from the University of Maryland (UMD) subreddit using the Reddit API and the Async PRAW (Python Reddit API Wrapper) library. The goal was to gather information related to academics and housing by analyzing posts and comments within the subreddit. We analyzed posts all the posts that the our code extracted using the API from `r/UMD` by searching for the following list of

keywords: ‘Housing’, ‘Dormitory’, ‘Residence hall’, ‘Apartment’, ‘Off-campus housing’, ‘Housing options’, ‘Roommate’, ‘Lease’, ‘Rent’, ‘Landlord’, ‘Housing cost’, ‘Housing amenities’, ‘Commute’, ‘Housing application’, ‘Housing lottery’, ‘Academics’, ‘Classes’, ‘Courses’, ‘Majors’, ‘Minors’, ‘Degree requirements’, ‘Class schedule’, ‘Professors’, ‘Lectures’, ‘Assignments’, ‘Exams’, ‘GPA’, ‘Academic advising’, ‘Study groups’, ‘Academic resources’, ‘Research opportunities’, ‘Internships’.

We created a script on [Google Colab](#) that would read the uploaded the data from the subreddit which was downloaded as a `.txt` file so that the program could extract the information and present it in a webpage. [The Github repository](#) hosts the webpage for the information presented with HTML. The information is displayed in cards for each posts with the post’s content, most upvoted comment, and most downvoted comment.

3.1.1 Insight

3.2 Survey

The second resource that we collected was the opinions of students; without that, there would not be much credibility in our project. To gather their opinions, we created a [survey](#) using Google Forms with the questions enumerated below:

1. Have you ever used Reddit when searching for information related to academics or housing (`r/UMD` is one example)?
2. How influential was it in your decision making?
3. Why Reddit (`r/UMD`)?
4. Approximately how much time did you spend searching for information?
5. If you’ve made a decision factoring in information found online, do you feel that your decision was correct?
6. Did you use any other sources besides Reddit?
7. If so, which ones? (write N/A if you selected "No" to the previous question)
8. Did you encounter conflicting information when browsing?
9. Would you recommend to a friend that Reddit or `r/UMD` is a good source for information?

The survey dataset included responses to questions about Reddit usage, decision-making experiences, perceived usefulness of Reddit content, and other relevant factors. Combining Reddit data with survey responses enabled us to analyze the relationship between online discussions and real-world decision-making outcomes. We collected about 40 responses

with the initial examination demonstrating that for the students that do use Reddit, they find its information useful, and also pair that information with other online sources like *PlanetTerp* and *records.umd.edu*. The results for each question are shown throughout the paper (Figures 1-4).

3.2.1 Insight

We discovered that lots of UMD students do in fact use Reddit for their information searching and its usage is best used when coupled with other resources like various UMD websites with course information or peers in order to vet sources when encountering conflicting information. And, of those who said they do not use Reddit for academic and housing information purposes, almost all reported they relied on personal connections with others and communicated face-to-face to gather information and make a corresponding decision. Many more survey responses are necessary, however, in order to produce statistical analyses of the data using various statistical tests like Chi-Squared and hypotheses tests.

Another insight we found while working on coming up with tools for gathering our dataset was related with the creation of the google form for conducting the survey. We aimed at creating the questions in such a way that they could be answered by everyone, this ensured that there was no bias in data collection and that we were covering all aspects. Upon further discussion we found that our survey could have been used with some improved designing of questions. For example, if someone did not use Reddit, then jumping directly to the questions related to other sources would have been beneficial. Also, perhaps categorizing the answers rather than permitting free-responses would ensure that qualitative analysis is more feasible and simple. But, this decision was also made to capture the widest range of responses and be the most representative of students' opinions. Another challenge was being able to gather enough responses. After leaving the survey open for several weeks hours, only 40 responses were gathered which is not sufficient to conclusively claim anything. If we were to continue this line of research, we would require significantly more responses and different methods of gathering them such as in-person asking around campus, speaking briefly to various computer science classes, and other methods.

4 Results

The core question our study seeks to address is how different types of information shared on Reddit, including comments, posts, and upvotes, influence students' decisions related to academic and housing goals. One key aspect we are exploring is the role of credibility and relevance in influencing decision-making. By understanding how students perceive and respond to different types of content on Reddit, we can gain insights

into the mechanisms by which online communities influence real-world decision-making processes.

Our analysis of the r/UMD subreddit has shown that Reddit is a pivotal source for UMD students, particularly in gathering insights on academic and housing decisions. This reliance is likely because highly upvoted comments and popular posts are perceived as more credible and informative. And, despite Reddit's popularity among students, the variability in content trustworthiness necessitates a critical evaluation approach. This is crucial considering Reddit's user-generated content where anonymity and varied user expertise can influence information accuracy. Reflecting on related works like "Social Signals that Drive Reddit Communities," it's evident that user karma, sentiment, and content relevance play significant roles in influencing the visibility and credibility of posts. These factors are instrumental in shaping students' perceptions and their subsequent decision-making processes. Using the Reddit API, we targeted and analyzed posts from the r/UMD subreddit related to student queries about academics and housing. The survey we conducted further substantiates our data extraction, indicating a strong correlation between Reddit discussions and student decisions in these areas. Based on our findings and the gaps identified in "Predicting Reddit Post Popularity Via Initial Commentary," future research should explore the mechanisms through which social signals affect user behavior on Reddit. Additionally, we should aim on developing educational tools to assist students in critically assessing online information's reliability. Lastly, machine learning techniques specified in previous papers would allow for predictions to be made about certain posts relating to the likelihood of their popularity.

5 Conclusion & Future Work

5.1 Conclusion

5.2 Future Work

Additional datasets on students' majors, genders, and grade levels with Reddit data may further offer a holistic approach to understanding the nuanced dynamics of online community engagement and decision-making among university students, so another form may be created that includes these changes and is more promoted to have an improved outreach. As a future step with respect to this research, by analyzing how these demographic factors intersect with Reddit usage patterns, we can uncover valuable insights into the factors shaping student behavior and decision-making processes. For example, exploring variations in Reddit engagement across different majors may reveal disciplinary preferences and academic interests among students, while examining gender differences in Reddit usage could shed light on potential disparities or biases within online communities. Additionally, investigating how students' grade levels influence their engagement with Red-

Have you ever used Reddit when searching for information related to academics or housing (r/UMD is one example)?

40 responses

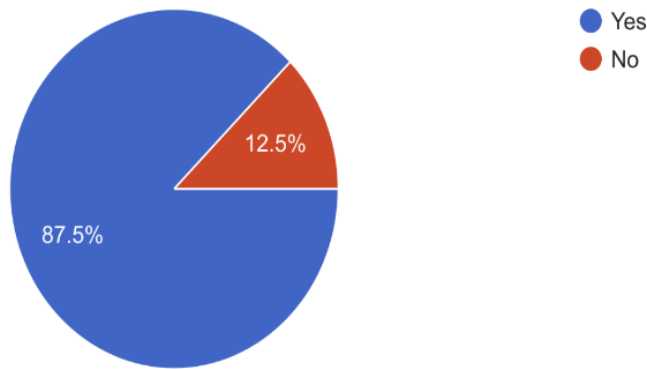


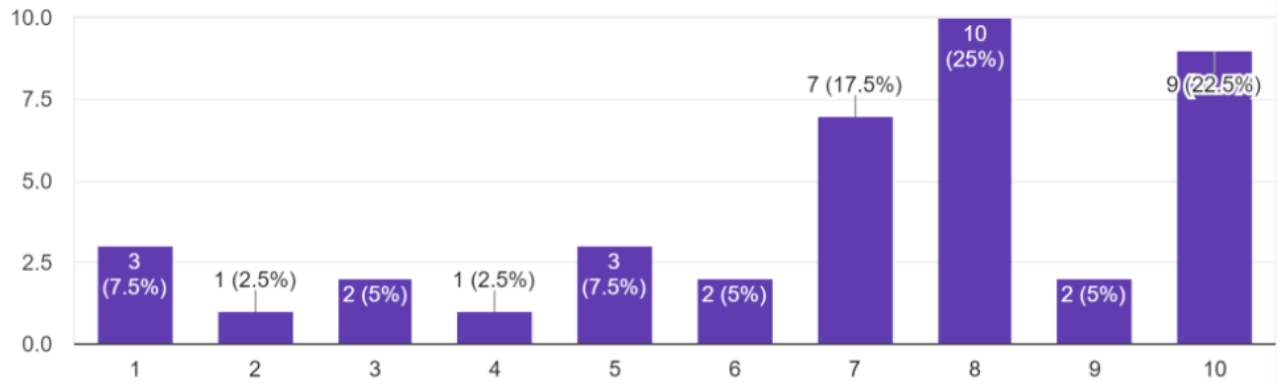
Figure 1: This graph demonstrates that a significant majority, 87.5%, of respondents used Reddit for academic and housing information. This strong usage underscores Reddit's role as a critical information source for UMD students.

dit and decision-making behaviors can provide insights into developmental trends and transitions in information-seeking and decision-making strategies over the course of students' academic journeys.

Moreover, creating machine learning models like the ones used in previous papers would allow us to predict with certainty *which* posts would become popular on the subreddit and thus the *decisions* students may make. The importance of this is that the school can be more accommodating for students' actions as they can plan in the future. For example, if a model predicts that one particular CMSC course seems to be gaining a lot of attention, either because it will be popular in the upcoming semester or because students are telling their peers to avoid the class, the university can react accordingly and increase the course capacity or revitalize the incentives to take the course, respectively.

How influential was it in your decision making?

40 responses



Approximately how much time did you spend searching for information (in hours) ?

40 responses

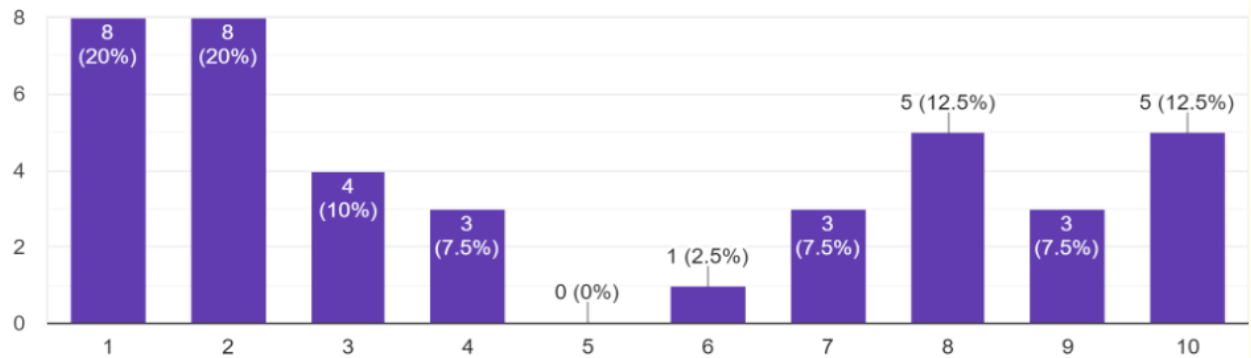
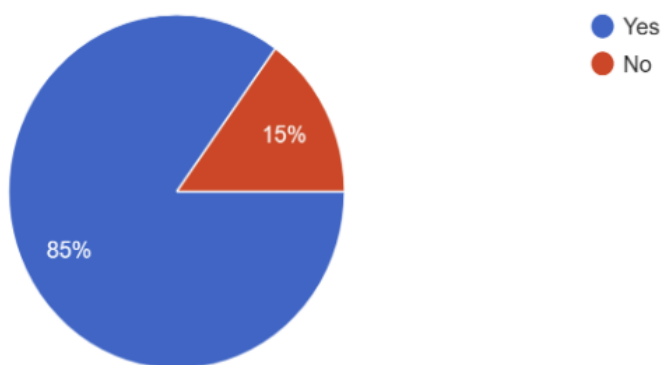


Figure 2: The influence of Reddit in decision-making varied, with some students rating it highly. Additionally, most students spent between 1 to 4 hours researching information, indicating moderate engagement with content to make informed decisions.

If you've made a decision factoring in information found online, do you feel that your decision was correct?

40 responses



Did you use any other sources besides Reddit?

40 responses

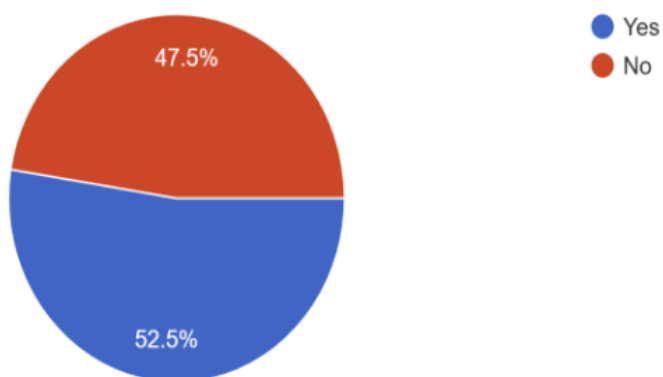
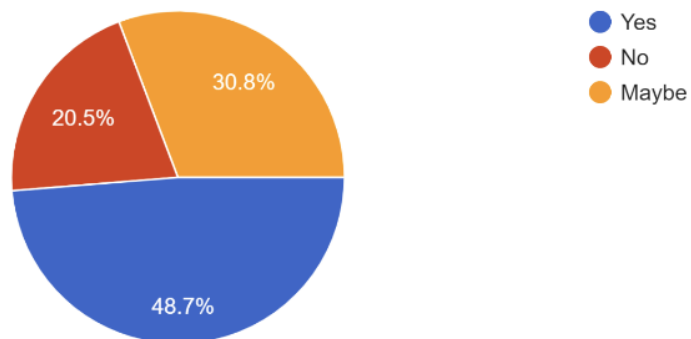


Figure 3: 85% felt their decisions based on Reddit information were correct, showing high satisfaction. Also, over half used other sources besides Reddit, highlighting a balanced approach to information gathering.

Did you encounter conflicting information when browsing?

39 responses



Would you recommend to a friend that Reddit or r/UMD is a good source for information?

40 responses

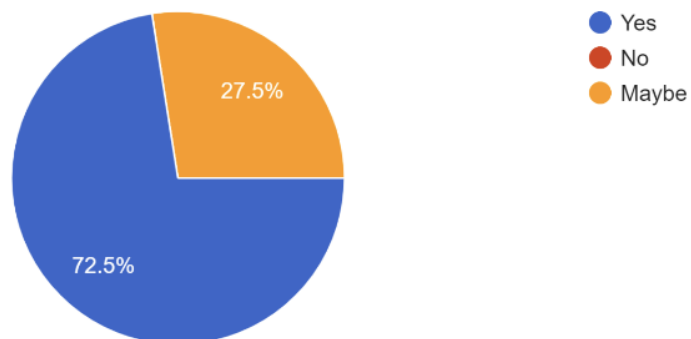


Figure 4: Nearly 70% encountered conflicting information, which may affect decision certainty. Despite this, 72.5% would still recommend Reddit as a reliable information source, reflecting trust in the community-driven advice despite some inconsistencies.