# Do top school students tend to criticize professors more?: Mining and Summarizing comments on RateMyProfessor.com

Ziqi Tang, Yutong Wang
University of Rochester
ztang14@ur.rochester.edu, ywang343@ur.rochester.edu

**ABSTRACT**

Student reviews and comments on RateMyProfessor.com reflect the most realistic learning experiences of each student. It provides a brand new point of view to examine the teaching quality of the lecturers. In this paper, we propose a framework of understanding these comments. First, we partition our data into different comparison groups. Next, we perform exploratory data analysis to delve into the data. Then we implement Latent Dirichlet Allocation and emotion analysis to extract topics and learn about emotion expressions of the comments. We reveal exciting insights about the characteristics of college students and professors. Furthermore, our study proves that student reviews and comments contain crucial information and could serve as one of the essential references for enrollment of courses or universities.

## INTRODUCTION

Since 1983, the U.S. News & World Report has been publishing rankings for the colleges and universities each fall. These rankings have remarkable impacts on applications, admissions, enrollment decisions, as well as pricing policies of tuition [11]. It is an important reference for not only students and parents, but also institutions and professors. The rankings methodology measures and calculates a variety of factors, and it has been continuously refined based on user feedback, discussions with institutions and education experts, literature reviews and their own data [12]. The current rankings methodology considers the following factors as measurements, and the indicator weights respectively are: Graduation and Retention Rates (22%), Undergraduate Academic (20%), Faculty resources (20%), Financial Resources (10%), student selectivity for entering class (7%), Graduation Rate performance (8%), Social Mobility (5%), Graduate Indebtedness (5%) and Alumni Giving Rate (3%). This measurement takes a good number of subjective factors into consideration. However, the learning experiences of students are objective and personally, which cannot be represented by the ranking scores. Therefore, a worldwide professor rating website, RateMyProfessors.com is a great resource to learn about the hidden knowledge which the U.S. News Rankings don't show.

Rate My Professor is a website that allows students to anonymously rate their professors and write comments and suggestions. The website claims that users have added more than 19 million ratings, 1.7 million professors and over 7,500 schools to the website, and there are more than 4 million college students each month using this website [1]. These text are a great resource to study the following topics: features of different universities, learning experiences of students; and course and lecture qualities. Past literature has primarily examined the usefulness and validity of these ratings [14]; and the correlation levels between easiness, clarity and helpfulness of lecturers [15]. Yet the data on Rate My Professors have more hidden information to discover. A unique feature of Rate My Professors is that it has professor reviews from different tiers of universities, such as Ivy League schools, Big Ten schools, and Community colleges. These reviews are all discussing an identical topic, which is the feelings and suggestions of taking a course from the universities. The uniqueness provides an opportunity to conduct a plausible control variable experiment to learn about the characteristics of students from different universities or colleges.

In particular, this study makes several contributions: In order to mine behaviors and characteristics of these groups, we perform the following methods:

1) Exploratory data analysis
2) Emotion analysis using a lexical approach
3) Several topic-modeling methods

on the following partitions:

1) Ivy League vs. Big Ten vs. Community college
2) High rating professors vs. low rating professors

### Collection and preprocessing

Rate My Professors data was scraped from the website. We selected about 75 universities based on the U.S. News college rankings 2020. The rationale of our selection was the following: 8 Ivy league schools represented the top ranked private universities; 10 Big Ten Academic Alliance Member universities represented the top ranked public universities; top 15 ranked community colleges in the United States represented the community colleges. Also, we selected the top 25 ranked universities and ranked 100 - 125 universities in the United States. For each university in our selections, we picked 60 the most-rated professors; and for each professor page, we scraped 20 the most recent comments. In total, we collected 87,436 data records, containing the following attributes: "Professor ID", "Professor Name", "University", "Department", "Course ID", "Quality score", "Difficulty score", "Comments". Each data record represented a review of a student on a course.

We partitioned the collected data into several datasets. The rationale was the following:

1) Based on the school type, we partitioned the data into three categories: Private vs public vs. community college. Namely, Ivy League vs. Big Ten vs. Community colleges.
2) Based on the average rating score of the professors. We calculated the average quality scores of each professor, and selected those professors with average scores higher than 4.0 and lower than 2.0, (score ranged from 1.0 to 5.0), as the high rating professor and low rating professor group. Namely, high rating professors vs. low rating professors.
3) Based on the quality score of each comment. We also prepared datasets for good quality score comments and bad scores quality score comments. Three categories were comments with scores higher than 4.0, comments with scores lower than 2.0 and comments with scores in between. Namely, high rating comments vs. low rating comments vs. low rating comments.

In the end, we had 11 datasets, and three groups of comparison to do. Note that these datasets may overlap with each other.

### Analysis Result

#### Exploratory data analysis

- The word counts,shown in Figure 1, indicates that most of the comments had around 60 words. All groups have the identical distribution. The only difference is that Ivy League students use short phrases more often than other groups.

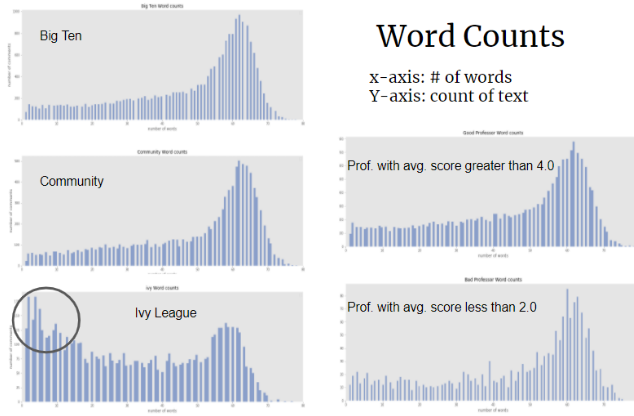


Figure 1

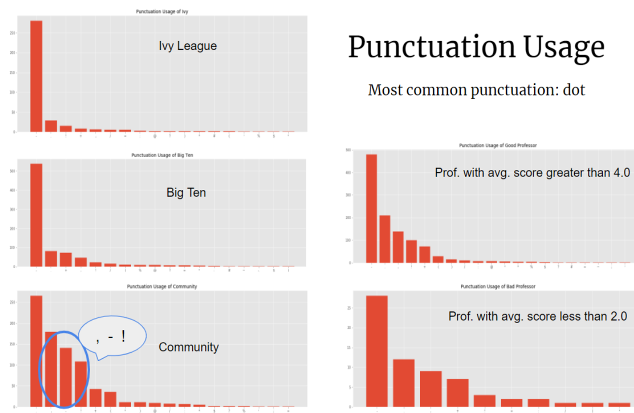


Figure 2

- The Punctuation usage, shown in Figure 2, demonstrates the most common used punctuation was dot.The distributions for all groups are similar as well. The only difference is that community college students use commas, dash and exclamation marks with a higher frequency than other groups.
- Figure 12 is a count plot of average quality ratings of all professors, and it has a skewed left distribution.
- Figure 13 is a count plot of quality ratings from all schools (75 schools)
- From the pie charts in Figure 14, the proportions of quality ratings of the three different groups of schools shows Community College students give more high (5/5) ratings, and Ivy League students give less low (1/5) ratings. Hence, this answers our initial questions. Top school students are not stricter when rating their professors and course quality.
- The pie charts in Figure 15 shows the proportions of difficulty ratings of the three different groups of schools, and the distributions are very similar.
- The correlations between quality ratings and diffi-

culty ratings for Ivy League, Big Ten and Community college are [-0.178, -0.424, -0.515], respectively. All groups have negative correlation values which implies the quality rating decreases when difficulty rating increases, or the opposite way. Ivy League's correlation is very close to zero which means there is nearly no relationship between quality ratings and difficulty ratings of Ivy League schools. Moreover, students from Big Ten schools and Community Colleges are more likely to give a higher rate when the course is easy.

## Quality Rating Analysis Using Topic Modeling

*BiGram vs. TriGram vs. Multi-Grain LDA*

In order to find out what factors influence the quality ratings, we performed Latent Dirichlet Allocation to extract topics of the comments. We implemented a few types of topic modeling methods: LDA, BiGram LDA and TriGram LDA using Gensim library. Also, we applied traditional LDA and Multi-Grain LDA using tomotopy library. Gensim is a well-known python library for topic modeling, and tomotopy is a new library that provides functions for topic modeling. Advantages of tomotopy are the capability of dealing with large scale datasets, significantly faster running time than Gensim (5 to 10 times faster than Gensim), and its availability for implementing Multi-Grain LDA model. Multi-Grain LDA takes both local topics and global topics into consideration when performing topic modeling. Therefore, we decided to examine the tomotopy Multi-Grain LDA model for our project. BiGram, TriGram and Multi-Grain LDA models are similar algorithms as traditional LDA. However, they have an additional step which adds N-gram phrases to increase the model's complexity, and this could be useful in boosting the model's performance. In our case, the BiGram model has phrases like: "easy-A", "office-hour", "online-course", etc. For the TriGram model, there are phrases like: "extra-credit-opportunity", "attendance_isn_mandatory", etc.

In order to evaluate the performance of all of these models, we used coherence score, pyLDAvis visualization, log-likelihood and manually checking method as our evaluation metrics. For LDA, BiGram LDA and TriGram LDA models using Gensim, their coherence score comparison plot is shown in Figure 4. Furthermore, we used pyLDAvis topic modeling visualization tool to analyze the performance of models. There is an example in Figure 3, the bubbles represent different topics. The optimal situation is all topic bubbles are spread in different quadrants which means they exhibit unique information instead of represent the same topics. For the Multi-Grain LDA model using tomotopy, the library does not generate coherence scores which was a downside of this library. Therefore, we decided to manually check all the topics these models generated and choose which one made more sense to us for the rest of the project. Figure 5 shows the result topics we created using BiGram LDA, TriGram LDA and Multi-Grain LDA methods. They were all generated from the same dataset (Community college lower quality rating comments) and had the same number of top topics selected (nine topics). A major portion of the topics are similar. However, TriGram LDA model covered the most of the topics. For instance, we saw key word "online" from the result of TriGram LDA, and since this was a Community College dataset, Community Colleges tend to have more online classes than other groups, and that could be a factor that students consider when they rate the course quality. Moreover, we also saw "accent" for the first time from the result of TriGram LDA, and this is a very interesting factor to include because many students actually have a different time understanding their professors' accents. The communication experience is an important aspect of course quality rating.

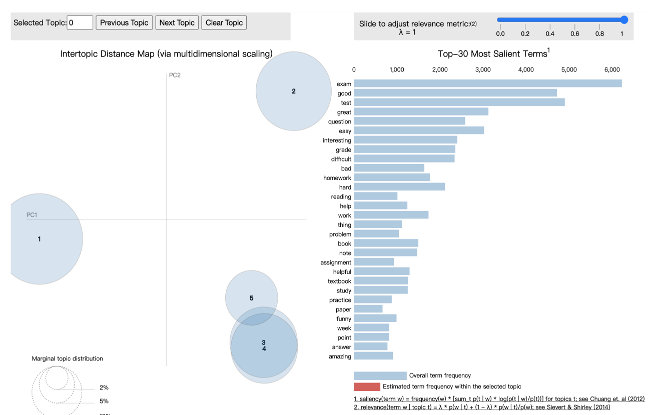


Figure 3

*Ivy League VS. Big Ten VS. Community Colleges*

*Higher Ratings (5-4):* The key words of topics of higher quality ratings of three partition groups are listed in figure 6. There are lots of factors that students mentioned in the comments when giving higher ratings. For example, school works(homework, test, exam), extra help(office_hour), professor's personality(friendly, humor, entertaining) and etc. Meanwhile, some unexpected words stand out in the table: "tough", "boring", "strict", and this imply that these are not negatively affecting Ivy League and Community College's quality ratings. Also,

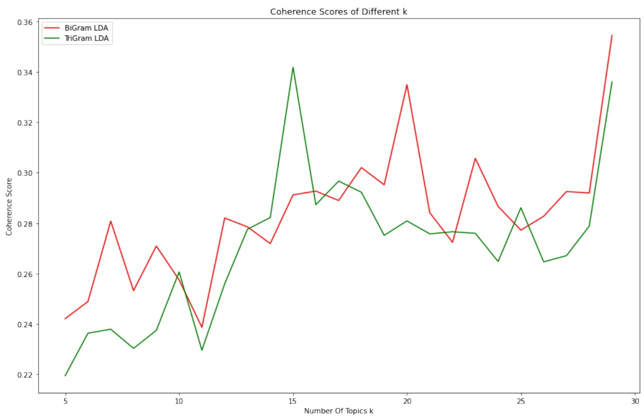


Figure 4

both Big Ten and Community College students mention "extra_credit", "grade" more often under this partition. The word "friend" appears in Big Ten's topics, and this might imply students in Big Ten schools are more likely to get along with their professors like friends.

*Lower Ratings (2-1):* The key words of topics of lower quality ratings of three partition groups are listed in figure 7. There are lots of factors that students mentioned in the comments when giving lower ratings. For example, school works(homework, test, exam), organization of the content(unclear, disorganized, useless), professor's attitude(manner, rude, arrogant) and etc. One thing to point out is that "cost" is a common factor through all schools, so the cost of textbooks, supplies and software has significant negative effects on quality ratings.

*Middle Ratings (4-2):* The key words of topics of middle quality ratings of three partition groups are listed in figure 8. Since middle rating comments are usually not too extreme, there is not a lot of new discovery. Nonetheless, "accent" appears under Big Ten's topics, and it is in Community College's topics for lower ratings. This suggests that Big Ten school students may have a higher tolerance on professors' accents than Community College students.

### *High Rating Professors vs. Low rating professors*

The key words of topics of comments of professors with average quality rating higher than 4 and lower than 2 are listed in figure 9. One thing to notice is the coherence score of higher average rating professors is lower which means the topics of these comments are more dissimilar. Factors that affect the average ratings of professors are: grade, difficulty, organization of the contents, personality, extra help, passion, knowledge, fairness etc. "voice" and "recitation" appear under lower average rating professors category, and this is the only time they appear. This implies communication is critical to students' experience in classes, and also professors teaching science classes(Physics, Chem, Bio) that have recitation sections tend to have lower average ratings.

## Emotion analysis using a Lexical Approach

The LIWC2015 program is a toolkit which includes the main text analysis module along with a group of predefined internal lexicons. The text analysis module compares each word in the text against the dictionary and then identifies which words are associated with which psychologically-relevant categories [16]. It has been used on previous studies for sentiment analysis on text data from social media [6]. LIWC2015 provides about a hundred of psychologically-relevant categories, we selected around 20 categories for our analysis.

### *Ivy League vs. Big Ten vs. Community college*

After we obtained LIWC scores for each data record, we calculated the average scores and standard deviations. Figure 10 shows the LIWC results for our first comparison group. Some interesting categories stood out: positive emotion, anxiety, achievement, sexual, and gender. We ran t-test on these categories and LIWC grand average scores. The two-tailed P values for Position emotion, Achievement and Male reference were all less than 0.001. ($P < 0.001$). By conventional criteria, the differences were considered to be statistically significant.

1) The positive emotion scores for college students were overall higher than the average. The Ivy League students score was not only higher than the grand average, but also higher than the other groups. It indicates that students from top ranked private schools do not tend to criticize professors more, instead, they praise the professors more often than other groups.
2) The Achievement score for community college students was higher than other groups. Our interpretation was that the community college students may had jobs previously, and they decided to attend community college because they wanted to receive more education and learn more skills. They possibly have clearer motivation and goals than other groups. Therefore they tend to talk about achievement-related topics more often in their comments.
3) The female reference score for community college and male reference score for Ivy League schools stood out. The gender reference scores were measured when the students mentioned gender related phrases, such as he or she. Due to the fact that Rate My Professor website does not record the

| Group | Topics | Top topic words |
|---|---|---|
| MultiGrin LDA | exam, test, assignment, **communication skills**, homework, personality, grader | 1.**book, assignment,** thing, college, answer, **instructions,** entire, teaching, boring, summer<br>2. day, clear, study, super, essay, **English, unclear, information,** weeks, impossible<br>3. semester, tough, terrible, awful, **unorganized,** nice, word, better, reason<br>4. horrible, question, **grades, extra, feedback,** helpful, answers, lots, videos, essays<br>5. work, things, wrong, **paper**, hours, multiple, end, **exam**, worse, attitude'<br>6. notes, **quizzes**, credit, points, little, high, **heavy**, best, major, page'<br>7. exams, week, **person**, gpa, sure, review, care, quiz, **avoid**, luck'<br>8. difficult, rude, papers, textbook, different, **reviews**, problems, times, grading, lab'<br>9. final, **grader**, easy, help, life, exam, topic, stupid, board, didnt' |
| BiGram LDA | exam,test,assignment, homework, personality, grader, **online experience** | 1.test, grade, bad, question, **assignment**, work, good, **homework**, hard, book<br>**2.exam**, thing, **online**, **final**, difficult, note, recommend, point, essay, easy,<br>3.study, tough, pass, credit, long, job, **study guide**, understanding, understand, attendance<br>4. paper, quiz, **topic**, **major**, term, dumb, correct, new, bit, favor<br>5. chapter, fact, **technology**, use, **content**, end, stuff, memorize, second, important<br>6. rude, semester, person, teaching, awful, mistake, time, ready, **response**, heavy<br>7. problem, **reading**, well, video, focus, discussion, regret, post, stress, concept<br>8. talk, **feedback**, fast, effort, funny, error, sarcastic, call, classmate, period<br>9. life, nice, **grader**, terrible, boring, joke, college, direction, unclear, fail |
| TriGram LDA | exam,test,assignment, **communication skills**, homework, personality, grader, **online experience** | **1.test**, **grade**, bad, question, **assignment**, work, good, homework, thing, rude<br>2. **exam**, hard, **online**, book, horrible, difficult, recommend, point, **review**, hour<br>**3. instruction**, **reading**, reason, video, focus, **discussion**, post, bit, phone, place<br>4. teaching, **essay**, life, nice, talk, page, tough, **accent**, dumb, attendance<br>5. problem, study, chapter, credit, care, impossible, overall, stuff, avoid, **explanation**<br>**6. paper**, easy, mistake, pass, **feedback**, **topic**, use, job, heavy, cost<br>7. waste, term, step, speech, favor, opinion, campus, tutor, **struggle**, follow<br>**8. content**, fast, **attitude**, super, effort, error, act, classmate, demand, aware<br>9. little, high, time, direction, **grader**, unclear, board, fun, response, comment |

Figure 5

| High | Ivy League | Big Ten | Community Colleges |
|---|---|---|---|
| TriGram LDA Topics Key words | exam, test, homework<br>approachable, friendly, humor<br>thank, grateful<br>office hour, helpful<br>experience<br>effort<br>**tough boring**<br>passion | exam, test, homework<br>**extra_credit, grade**<br>helpful<br>thank, grateful<br>entertaining, enjoyable<br>opportunities<br>office_hour<br>passion<br>**friend** | exam, test, homework<br>**extra_credit, grade**<br>nice caring<br>humor joke<br>opportunities<br>effort<br>participation<br>**stress lazy**<br>**strict tough** |
| Coherence | 0.469 | 0.388 | 0.433 |

Figure 6

gender of the professor, and we collected a fixed number of comments from each professor, this score generated from gender reference words is the most reliable way to count the number of comments for male professors and female professors. Our analysis indicated that there are more male lecturers in Ivy league schools and more female lecturers in the community colleges. Our interpretation was that for research-based institutions, like Ivy League schools and Big Ten schools, the professors and lecturers are required to constantly work on projects for long term. Community colleges, on the other hand, have more part-time lecturer positions. For female researchers who might have to take care of family and children at the same time, teaching part time at community college seems a reasonable decision.

4) The anxiety scores were considered to be statistically insignificant. Based on some literature, our expectation was that students attending top ranked private colleges have a higher chance to feel depression and pressure [7]. However, the LIWC result showed that the students did not express pressure and anxiety in their reviews. Our interpretation was that these comments were mostly written after the final exams or projects. The students no longer feel anxious at the time they post the comments.
5) The sexual scores were considered to be statistically insignificant. The sexual category contains phrases that describe the appearance of the professors. This could indicate whether the appearance could affect the student ratings and comments. Our study showed there is no evidence to prove the existence of connection between appearance and student ratings.

| Low | Ivy League | Big Ten | Community Colleges |
|---|---|---|---|
| TriGram<br>LDA<br>Topics<br>Key words | exam, test, homework<br>grade<br>disorganized, unclear<br>useless, unhelpful<br>hard, difficult<br>**manner, rude, arrogant**<br>feedback<br>**cost** | exam, test, homework<br>grad<br>office hour<br>unclear, unseless, irrelevant,<br>confusing<br>hard, difficult<br>rude<br>feedback<br>**cost** | exam, test, homework<br>grade<br>unclear<br>**accent**<br>dumb<br>hard tough<br>struggle<br>feedback<br>**cost** |
| Coherence | 0.558 | 0.446 | 0.495 |

Figure 7

| Middle | Ivy League | Big Ten | Community Colleges |
|---|---|---|---|
| TriGram<br>LDA<br>Topics<br>Key words | exam, test<br>disorganized<br>passion, enthusiastic<br>boring, tough<br>personality, outgoing | exam, test<br>grade curve<br>unclear confusing<br>**accent**<br>passion, enthusiastic<br>annoying | exam, test<br>grade<br>fun simple<br>fair<br>opportunity<br>helpful<br>fair<br>strict<br>passion |
| Coherence | 0.540 | 0.435 | 0.493 |

Figure 8

| | Average Rating > 4.0 | Average Rating < 2.0 |
|---|---|---|
| TriGram<br>LDA<br>Topics<br>Key words | grade, extra credit, pass<br>simple, easy<br>hard, tough<br>care<br>clear, prepared<br>engaging<br>interesting, enjoyable<br>helpful, patient, approachable, feedback<br>knowledgeable, passion<br>fair, honest | grade, GPA<br>difficult, hard, tough<br>boring<br>unstructured<br>condescending, rude<br>vague, pointless, unclear<br>drop<br>instruction<br>**voice**<br>**recitation** |
| Coherence | 0.333 | 0.504 |

Figure 9

*High Rating Professors vs. Low rating professors*

After we obtained LIWC scores for each data record, we calculated the average scores and standard deviations. Figure 11 shows the LIWC results for our first comparison group. Some interesting categories stood out: achievement and gender. We ran t-test on these categories and LIWC grand average scores. The two-tailed P value for Achievement and gender reference were both less than 0.001. ($P < 0.01$). By conventional criteria, the differences were considered to be statistically significant.

1) The achievement score for high rating professors is higher than the low rating professors. This may indicate that, beside the general impressions people have for a good professor or a lecturer, students think a good lecturer also needs to know how to motivate the students.
2) The female reference score for low rating professors is higher, and the male reference score for high rating professors is higher. This shows that there are more low rating female professors and more high rating male professors. It may imply that students are more strict on female lecturers than male lecturers.

## CONCLUSION AND FUTURE WORK

In this paper, we have presented a framework of evaluating learning experiences of college students from a more objective perspective. We first partitioned the scraped data into different groups, and then applied several LDA models to understand topics of the comments. Furthermore, we performed emotion analysis using LIWC2015. We discovered a few interesting findings and approved Rate My Professors contains information that may be helpful for people, especially for students, scholars, professors and institutions. There are three possible works for the future.Firstly, we could propose a good partition strategy to divide the data by departments, subjects or courses. Then run the same procedures to discover knowledge. Secondly, we could keep tracking the comments at different time points. Our current datasets mostly contain comments from 2018 to 2020, and most

| Category | Example | Ivy | Big Ten | Community | LIWC Twitter Mean | LIWC Grand Mean (SD) |
|---|---|---|---|---|---|---|
| **Psychological Processes** | | | | | | |
| Positive emotion | Love, nice | 11.53 | 6.77 | 6.84 | 5.48 | 3.67(1.63) |
| Negative emotion | Hurt, ugly | 2.54 | 2.28 | 1.67 | 2.14 | 1.84(1.09) |
| Anxiety | Worried, fearful | 0.34 | 0.32 | 0.28 | 0.24 | 0.31(0.32) |
| Anger | Hate, kill | 0.51 | 0.39 | 0.28 | 0.75 | 0.54(0.59) |
| Sadness | Crying, sad | 0.28 | 2.78 | 3.57 | 0.29 | 0.41(0.40) |
| **Drives** | | | | | | |
| Achievement | Win, success | 2.38 | 1.68 | 12.99 | 1.82 | 1.30(0.82) |
| Reward | Benefit, prize | 4.70 | 3.15 | 9.21 | 1.07 | 1.46(0.81) |
| **Biological processes** | | | | | | |
| Sexual | Sexy, charming | 0.14 | 0.10 | 0.08 | 0.24 | 0.13(0.30) |
| **Time Orientations** | | | | | | |
| Past Focus | Ago, did | 3.07 | 3.01 | 3.11 | 2.81 | 4.64(2.06) |
| Present focus | Today, now | 10.45 | 12.69 | 12.99 | 11.74 | 9.96(2.80) |
| Future focus | Will, soon | 2.58 | 1.18 | 1.45 | 1.6 | 1.42(0.90) |
| **Social Processes** | | | | | | |
| Female references | Her, mom | 1.44 | 1.36 | 2.23 | 0.54 | 0.98(1.26) |
| Male references | His, dad | 3.93 | 3.58 | 3.57 | 0.84 | 1.65(1.34) |
| **Informal language** | | | | | | |
| Swear words | ****, **** | 0.15 | 0.07 | 0.05 | 0.49 | 0.21(0.37) |
| Netspeak | Btw, lol | 0.26 | 0.23 | 0.27 | 3.23 | 0.97(1.17) |
| Assent | Agree, OK | 0.86 | 0.33 | 0.33 | 1.82 | 0.95(0.72) |
| Nonfluencies | Er, hm, umm | 0.35 | 0.29 | 0.19 | 0.39 | 0.54(0.49) |
| Fillers | I mean, you know | 0.005 | 0.01 | 0.009 | 0.04 | 0.11(0.27) |

Figure 10

Notes: Grand Mean are the unweighted means of the six genres; Mean SD refers to the unweighted mean of the standard deviations across the six genre categories.

of the comments are posted at two time points: December and May, which are the ends of spring and fall semesters. In other words, we have data in 5 different time points which is not enough to support us to analyze changes along time. With more data from different time points, we are able to study individual professor's teaching style changes and look at this problem from a brand new point of view. Thirdly, many in-person lectures are switched to online lectures due to COVID-19 and quarantine. A possible project idea is first determine the courses that are transformed from in-person to online and then learn the changes from the student's experiences.

## References


[1] Rate my professors about page.

[2] Muhammad Abulaish, Jahiruddin, Mohammad Najmud Doja, and Tanvir Ahmad. Feature and opinion mining for customer review summarization. In Santanu Chaudhury, Sushmita Mitra, C. A. Murthy, P. S. Sastry, and Sankar K. Pal, editors, *Pattern Recognition and Machine Intelligence*, pages 219–224, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[3] J.B. Arbaugh. How instructor immediacy behaviors affect student satisfaction and learning in web-based courses. *Business Communication Quarterly*, 64(4):42–54, 2001.

[4] David Blei, Andrew Ng, and Michael Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 05 2003.

[5] David M. Blei and John D. Lafferty. Dynamic topic models. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, page 113–120, New York, NY, USA, 2006. Association for Computing Machinery.

[6] Long Chen, Hanjia Lyu, Tongyu Yang, Yu Wang, and Jiebo Luo. In the eyes of the beholder: Analyzing social media use of neutral and controversial terms for covid-19, 2020.

[7] William Deresiewicz. Don't send your kid to the ivy league, 2014.

[8] Oren Etzioni, Michael Cafarella, Doug Downey, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S. Weld, and Alexander Yates. Unsupervised named-entity extraction from the web: An experimental study. *Artif. Intell.*, 165(1):91–134, June 2005.

[9] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. KDD '04, page 168–177, New York, NY, USA, 2004. Association for Computing Machinery.

[10] George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller. Introduction to WordNet: An On-line Lexical Database*. *International Journal of Lexicography*, 3(4):235–244, 12 1990.

| Category | Example | Rating >= 4.0 | Rating <= 2.0 | LIWC Twitter Mean | LIWC Grand Mean (SD) |
|---|---|---|---|---|---|
| **Psychological Processes** | | | | | |
| Positive emotion | Love, nice | 9.17 | 3.67 | 5.48 | 3.67(1.63) |
| Negative emotion | Hurt, ugly | 1.30 | 4.84 | 2.14 | 1.84(1.09) |
| Anxiety | Worried, fearful | 0.18 | 0.95 | 0.24 | 0.31(0.32) |
| Anger | Hate, kill | 0.22 | 0.80 | 0.75 | 0.54(0.59) |
| Sadness | Crying, sad | 0.25 | 0.46 | 0.29 | 0.41(0.40) |
| **Drives** | | | | | |
| Achievement | Win, success | 2.38 | 1.68 | 1.82 | 1.30(0.82) |
| Reward | Benefit, prize | 4.70 | 3.15 | 1.07 | 1.46(0.81) |
| Risk | Doubt, danger | 0.47 | 1.82 | 0.46 | 0.47(0.41) |
| **Biological processes** | | | | | |
| Sexual | Sexy, charming | 0.12 | 0.03 | 0.24 | 0.13(0.30) |
| **Social Processes** | | | | | |
| Female references | Her, mom | 1.52 | 1.78 | 0.54 | 0.98(1.26) |
| Male references | His, dad | 3.89 | 3.36 | 0.84 | 1.65(1.34) |
| **Informal language** | | | | | |
| Swear words | ****, **** | 0.04 | 0.18 | 0.49 | 0.21(0.37) |
| Netspeak | Btw, lol | 0.24 | 0.27 | 3.23 | 0.97(1.17) |
| Assent | Agree, OK | 0.52 | 0.16 | 1.82 | 0.95(0.72) |
| Nonfluencies | Er, hm, umm | 0.27 | 0.22 | 0.39 | 0.54(0.49) |
| Fillers | I mean, you know | 0.008 | 0.01 | 0.04 | 0.11(0.27) |

Figure 11

Notes: Grand Mean are the unweighted means of the six genres; Mean SD refers to the unweighted mean of the standard deviations across the six genre categories.

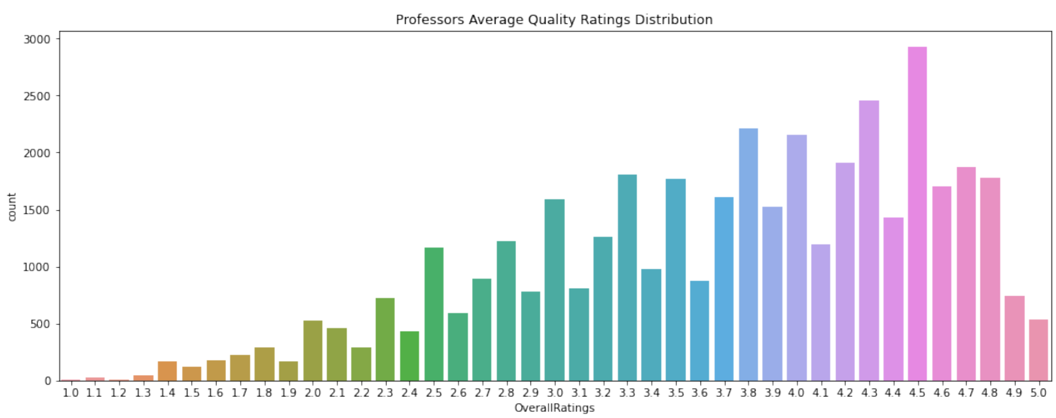


Figure 12

[11] James Monks and Ronald G Ehrenberg. The impact of us news and world report college rankings on admission outcomes and pricing decisions at selective private institutions. Working Paper 7227, National Bureau of Economic Research, July 1999.

[12] Robert Morse and Eric Brooks. How u.s. news calculated the 2021 best colleges rankings, Sep 2020.

[13] David Newman, Arthur Asuncion, Padhraic Smyth, and Max Welling. Distributed algorithms for topic models. *J. Mach. Learn. Res.*, 10:1801–1828, December 2009.

[14] James Otto, Douglas A. Sanford Jr, and Douglas N. Ross. Does ratemyprofessor.com really rate my professor? *Assessment & Evaluation in Higher Education*, 33(4):355–368, 2008.

[15] James Otto, Douglas Sanford, and William Wagner. Analysis of online student ratings of university faculty. *Journal of College Teaching Learning (TLC)*, 2:25–30, 01 2011.

[16] James Pennebaker, Ryan Boyd, Kayla Jordan, and Kate Blackburn. The development and psychometric properties of liwc2015. *University of Texas Libraries*, 2015.

[17] Kathleen M. Silva, Francisco J. Silva, Megan A. Quinn, Jill N. Draper, Kimberly R. Cover, and Alison A. Munoff. Rate my professor: Online evaluations of psychology instructors. *Teaching of Psychology*, 35(2):71–80, 2008.

[18] Ivan Titov and Ryan McDonald. Modeling online reviews with multi-grain topic models, 2008.

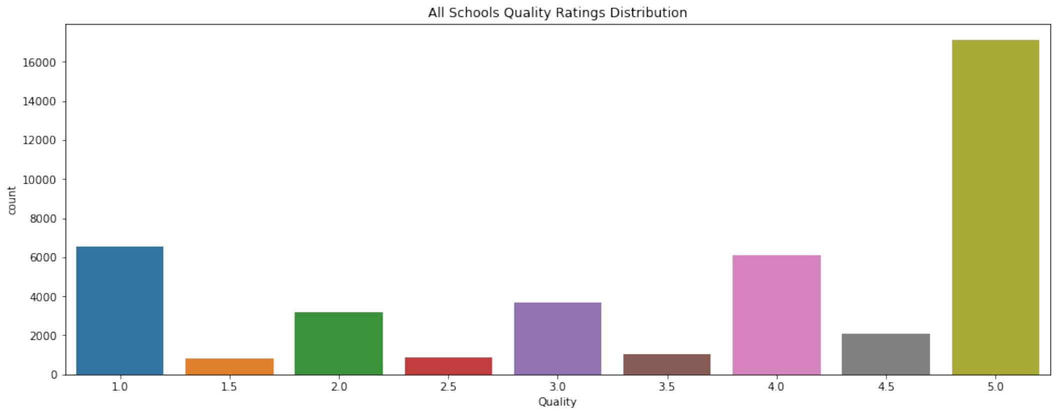


Figure 13

# Quality Rating Pie Charts

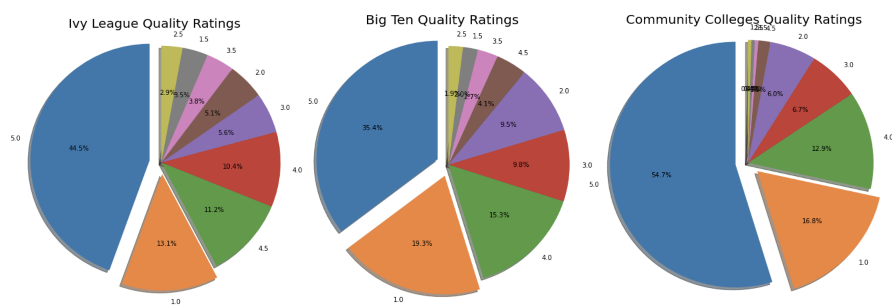


Figure 14

# Difficulty Rating Pie Charts

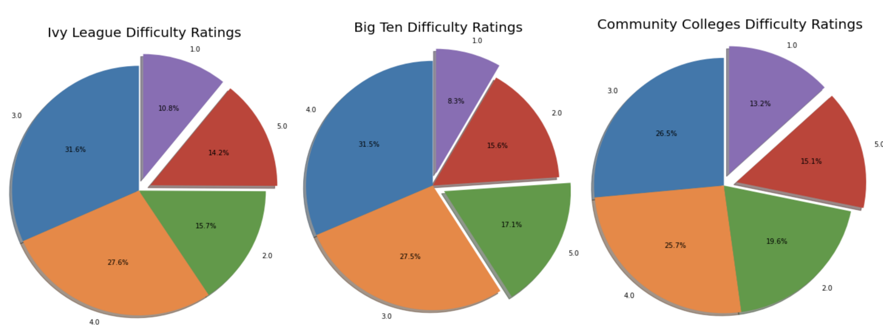


Figure 15