Are Top School Students More Critical of Their Professors? Mining Comments on RateMyProfessors.com

Ziqi Tang[1,2], Yutong Wang[1,2], and Jiebo Luo[1,3]

[1] University of Rochester
[2] {ztang14, ywang343}@ur.rochester.edu
[3] jluo@cs.rochester.edu

Abstract. Student reviews and comments on RateMyProfessors.com reflect realistic learning experiences of students. Such information provides a large-scale data source to examine the teaching quality of the lecturers. In this paper, we propose an in-depth analysis of these comments. First, we partition our data into different comparison groups. Next, we employ Latent Dirichlet Allocation and sentiment analysis to extract topics and understand the sentiments associated with the comments. We uncover interesting insights about the characteristics of both college students and professors. Our study confirms that student reviews and comments contain crucial information and can serve as informative references for enrollment in courses and universities.

Keywords: Topic Modeling · Emotion Analysis · RateMyProfessors.com.

1 Introduction

Since 1983, the U.S. News & World Report has been publishing rankings for the colleges and universities in the United States each fall. These rankings have remarkable impacts on applications, admissions, enrollment decisions, as well as tuition pricing policies [10]. It is an important reference for not only students and parents, but also institutions and professors. The ranking methodology measures and calculates a variety of factors, and has been continuously refined over time based on user feedback, discussions with institutions and education experts, literature reviews and their own data [11]. The current ranking methodology considers the following factors, along with indicator weights: 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 objective factors into consideration. However, the learning experiences of students are subjective and personal, which cannot readily be represented by the ranking scores. In this regard, a professor rating website such as RateMyProfessors.com is a great resource to uncover the hidden knowledge about the learning experience that the U.S. News Rankings can not account for.

RateMyProfessors.com is a website that allows students to anonymously rate their professors and write comments. 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 using this website each month [1]. Such massive text data is 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 [13], and the correlation levels between easiness, clarity and helpfulness of lecturers [14]. Yet the rich data on Rate My Professor contain more hidden information to discover. A unique feature of Rate My Professor is that it has professor reviews from different tiers of universities, such as Ivy League schools, Big Ten schools, and community colleges. These reviews discuss the same topic, which is the experiences of taking a course from a college professor. This provides an opportunity to conduct a plausible control variable experiment to learn about the characteristics of students and professors in different universities or colleges.

In summary, this study makes several contributions:

1. We conduct a large-scale study of the course learning experiences across the broad spectrum of universities and colleges in the United States.
2. We employ exploratory data analysis, topic modeling, and sentiment analysis to mine the behaviors and characteristics of different segments of colleges, students and professors.
3. We uncover interesting and useful insights that can help understand and improve college learning experience.

2   Data Collection and Preprocessing

RateMyProfessors data is scraped from the website and is publicly available in Github (https:/github.com/yutongwang901/RateMyProfessors). We select about 75 universities based on the U.S. News college rankings of 2020. The rationale of our selection is the following: The eight Ivy League schools represent the top ranked private universities, ten Big Ten Academic Alliance Member universities represent the top ranked public universities, and the top 15 ranked community colleges in the United States represent the community colleges. In addition, we select the top 25 ranked universities and those ranked in [100 - 125] in the United States. When ranking colleges, the admission, facilities & amenities statistics are the most important factors, even though they may not measure the subjective learning experience of each student. The comparison groups selected above are designed to be groups with different levels of average statistics. For instance, in 2020, the average undergraduate tuition of the Ivy League and Big Ten schools are $56,425$ and $33,716$, respectively (non-resident for public schools), while the average acceptance rates are 7% and 56%, respectively, and the 75th percentile SAT scores are 1565 and 1323, respectively. The above statistics indicate that there are clear differences between our comparison groups. In our study, we will examine whether these differences still hold in terms of learning experiences and student sentiments.

For each university in our selections, we picked the 60 most-rated professors (not the highest rated), and for each professor page, we scraped the most recent 20 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 represents a review by a student on a course.

We partitioned the collected data into several datasets (11 datasets for three types of comparison. Note that these datasets may overlap with each other):

1. Based on the school type, we partition the data into three categories: private (Ivy League), public (Big Ten), and community colleges.
2. Based on the average rating scores of the professors, we select those above 4.0 and below 2.0 (the full score range is from 1.0 to 5.0), as the high-rating professor and low-rating professor groups, respectively.
3. Based on the quality score of each comment, we create datasets for three categories: comments with a score above 4.0, below 2.0, and in between.

3   Quality Rating Analysis Using Topic Modeling

In order to find out what factors influence the quality ratings, we perform Latent Dirichlet Allocation (LDA) to extract topics of the comments. We implement a few types of topic modeling methods: LDA, BiGram LDA and TriGram LDA using the Gensim library. Also, we apply traditional LDA and Multi-Grain LDA using the Tomotopy library. The 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. Multi-Grain LDA takes both local topics and global topics into consideration when performing topic modeling. Therefore, we decide to examine the Tomotopy Multi-Grain LDA model for our study. BiGram, TriGram and Multi-Grain LDA models are similar algorithms to the traditional LDA. However, they have an additional step that adds N-gram phrases to increase the model's complexity, which 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 use coherence score, pyLDAvis visualization, log-likelihood, and manual checking as our evaluation metrics. For LDA, BiGram LDA and TriGram LDA models using Gensim, their coherence score comparison is shown in Figure 1. Furthermore, we use the pyLDAvis topic modeling visualization tool to analyze the performance of models. For the Multi-Grain LDA model using Tomotopy, the library does not generate coherence scores, which is a downside of this library. Therefore, we decide to manually check all the topics these models generate and choose the one that makes more sense to us. Figure 2 shows the resulting topics we create using BiGram LDA, TriGram LDA and Multi-Grain LDA methods. They are all generated from the same dataset (community college lower quality rating comments)

and have the same number of top topics selected (nine topics). A major portion of the topics are similar. However, the TriGram LDA model covers the most of the topics. For instance, we see key word "online" from the result of TriGram LDA. Since this is a community college dataset, we can infer that community colleges tend to offer more online classes than other groups, which could be a factor that students consider when they rate the course quality. Moreover, we also see "accent" for the first time from the result of TriGram LDA. 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.
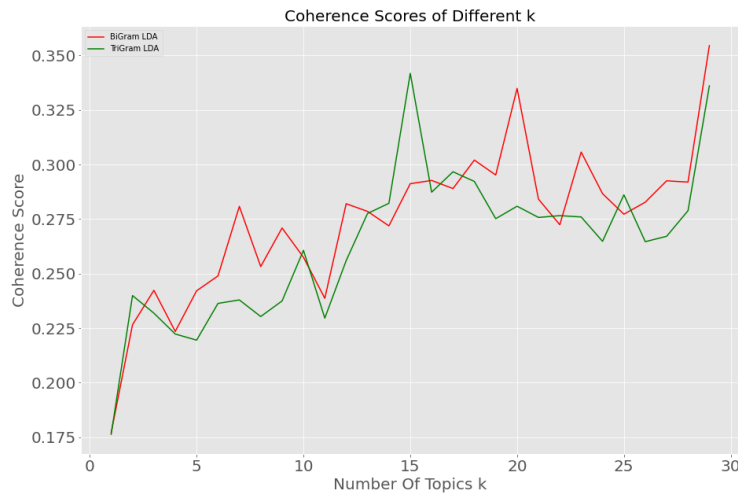

Fig. 1. Coherence score comparison between BiGram LDA and TriGram LDA models.

3.1   Ivy League vs. Big Ten vs. Community Colleges

High Ratings (4-5) The key words of the topics of high quality ratings for the three groups are listed in Figure 3. There are many factors that students mentioned in the comments when giving high ratings. For example, school works (homework, test, exam), extra help (office_hour), professor's personality (friendly, humor, entertaining), and so on. Meanwhile, some unexpected words stand out in the table: "tough", "boring", "strict", implying that these are not negatively affecting Ivy League and community college's quality ratings. In addition, both Big Ten and community college students mention "extra_credit", "grade" more often. The word "friend" appears in Big Ten's topics, implying students in Big Ten are more likely to get along with their professors like friends.

Low Ratings (1-2) The key words of the topics of low quality ratings for the three groups are listed in Figure 4. A number of factors are mentioned by students

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

Fig. 2. Comparison of topics from MultiGrin, BiGram and TriGram LDA models.

in the comments with low ratings. For example, school works (homework, test, exam), organization of the content (unclear, disorganized, useless), professor's attitude (manner, rude, arrogant), and so on. One thing to point out is that "cost" is a common factor through all schools as the cost of textbooks, supplies and software has significantly negative effects on quality ratings.

Middle Ratings (2-4) The topic key words of middle quality ratings are listed in Figure 5. The middle rating comments are usually not too extreme. We note that "accent" appears under Big Ten's topics, and in community college's topics for low ratings. This suggests that Big Ten school students have a higher tolerance for professors' accents than community college students.

3.2 High-Rating Professors vs. Low-Rating Professors

The key words in the comments for professors with an average quality rating higher than 4 and lower than 2 are listed in Figure 6. One thing to notice is

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

Fig. 3. Topic key words of higher ratings (4-5) of the three college groups.

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

Fig. 4. Topic key words of lower ratings (1-2) of the three college groups.

that the coherence score of higher 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, and so on. In contrast, "voice" and "recitation" appear in the lower rating professors category, and this is the only time they appear. This implies communication is critical to students' experience in classes, and professors teaching science classes (Physics, Chemistry, Biology) that have recitation sections tend to get low average ratings.

4   Sentiment Analysis Using A Lexical Approach

The LIWC2015 toolkit 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 [15]. It has been used on previous studies for sentiment analysis on text data from social media (e.g., [6]. LIWC2015 provides about a hundred psychologically-relevant categories, from which we select around 20 categories for our analysis.

Ivy League vs. Big Ten vs. Community Colleges After we obtain the LIWC scores for each data record, we calculate the average scores and standard deviations. Figure 7 shows the LIWC results for our first comparison group (Group A). Some interesting categories stand out: positive emotion, anxiety,

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

Fig. 5. Topic key words of middle ratings (2-4) of the three college groups.

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

Fig. 6. Topic key words of the two professor groups.

achievement, sexual, and gender. We run the t-test on these categories and LIWC grand average scores. The two-tailed P values for Positive Emotion, Achievement and Male Reference were all below 0.001 ($P < 0.001$). By conventional criteria, the differences are considered to be statistically significant.

We make the following observations:

1. The positive emotion scores for college students are overall higher than the average. The Ivy League students score is 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 is higher than other groups. Our interpretation is that the community college students may have had jobs previously and they decide to attend community college because they want 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 colleges and Male Reference score for Ivy League schools stand out. The gender reference scores are measured when the students mention gender related phrases, such as he or she. Due to the fact that Rate My Professor website does not record the gender of the professor, we collect a fixed number of comments from each professor. The score generated from gender reference words is the most reliable way to infer male and female professors. Our analysis indicates that there are more male

Category	Example	High-rating Professors	Low-rating Professors	Ivy League	Big Ten	Community	LIWC Twitter Mean	LIWC Grand Mean (SD)
Psychological Processes								
Positive emotion	Love, nice	**9.17**	3.67	**11.53**	6.77	6.84	5.48	3.67(1.63)
Negative emotion	Hurt, ugly	1.30	**4.84**	**2.54**	2.28	1.67	2.14	1.84(1.09)
Anxiety	Worried, fearful	0.18	**0.95**	**0.34**	0.32	0.28	0.24	0.31(0.32)
Anger	Hate, kill	0.22	**0.80**	**0.51**	0.39	0.28	0.75	0.54(0.59)
Sadness	Crying, sad	0.25	**0.46**	0.28	2.78	**3.57**	0.29	0.41(0.40)
Drives								
Achievement	Win, success	**2.38**	1.68	2.38	1.68	**12.99**	1.82	1.30(0.82)
Reward	Benefit, prize	**4.70**	3.15	4.70	3.15	**9.21**	1.07	1.46(0.81)
Risk	Doubt, danger	0.47	**1.82**	N/A	N/A	N/A	0.46	0.47(0.41)
Biological processes								
Sexual	Sexy, charming	**0.12**	0.03	**0.14**	0.10	0.08	0.24	0.13(0.30)
Time Orientations								
Past Focus	Ago, did	N/A	N/A	3.07	3.01	**3.11**	2.81	4.64(2.06)
Present focus	Today, now	N/A	N/A	10.45	12.69	**12.99**	11.74	9.96(2.80)
Future focus	Will, soon	N/A	N/A	**2.58**	1.18	1.45	1.6	1.42(0.90)
Social Processes								
Female references	Her, mom	1.52	**1.78**	1.44	1.36	**2.23**	0.54	0.98(1.26)
Male references	His, dad	**3.89**	3.36	**3.93**	3.58	3.57	0.84	1.65(1.34)
Informal language								
Swear words	****, ****	0.04	**0.18**	**0.15**	0.07	0.05	0.49	0.21(0.37)
Netspeak	Btw, lol	0.24	**0.27**	0.26	0.23	**0.27**	3.23	0.97(1.17)
Assent	Agree, OK	**0.52**	0.16	**0.86**	0.33	0.33	1.82	0.95(0.72)
Nonfluencies	Er, hm, umm	**0.27**	0.22	**0.35**	0.29	0.19	0.39	0.54(0.49)
Fillers	I mean, you know	0.008	**0.01**	0.005	**0.01**	0.009	0.04	0.11(0.27)

Fig. 7. LIWC results of two comparison groups. Group A: professors with average ratings above 4.0 vs. below 2.0. Group B: Ivy League vs. Big Ten vs. community colleges. Grand Mean is the unweighted mean of the six genres, and Mean SD refers to the unweighted mean of the standard deviations across the six genre categories.

professors in Ivy league schools and more female lecturers in the community colleges. Our interpretation is that for research-oriented institutions, like Ivy League schools and Big Ten schools, the professors are required to perform research and teaching full-time. Community colleges, on the other hand, have more part-time lecturer positions. For female professors who might have to take care of family and children at the same time, teaching part-time at community college seems to be a good option.

4. The anxiety scores are considered to be statistically insignificant. Based on the literature, our expectation was that students attending top ranked private colleges have a higher likelihood to feel depression and pressure [**?**]. However, the LIWC results show that the students did not express pressure and anxiety in their reviews. Our interpretation is 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 are 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.

High-Rating Professors vs. Low-Rating Professors Similarly, after we obtain the LIWC scores for each data record, we calculated the average scores and standard deviations. Figure 7 also shows the LIWC results for our second comparison group (Group B). Some interesting categories stand out: Achievement and Gender. We run the 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. The specific findings are:

1. The Achievement score for high-rating professors is higher than the low-rating professors. This may indicate that apart from the general impressions people have for a good professor, students think a good professor also needs to know how to motivate the students.
2. The Female Reference score for low-rating professors is higher, while the Male Reference score for high-rating professors is higher. There are more low-rating female professors and more high-rating male professors, perhaps implying that students are more critical on female professors.

5   Conclusion and Future Work

In this paper, we have presented a study of the learning experiences of college students from a more subjective perspective. We first partition the scraped data from RateMyProfessor.com into different groups and apply several LDA models to understand the topics of the comments. Furthermore, we perform sentiment analysis using LIWC2015. We discover a number of findings that may be helpful for improving the college learning experience for all parties involved, including students, professors and administrators. For instance, the key topic word "entertaining & enjoyable" only appears in the Big Ten professor comments, and the key topic word "lazy" only appears in the community college professor comments. It indicates that even though every professor has his/her own teaching style, professors from different schools do share some features in common, e.g., Ivy League professors are "boring" and "arrogant", Big Ten professors are "entertaining" but "confusing", while community college professors are "nice caring" and talk with "accent". Therefore, our findings can potentially guide students and parents to choose universities that not only match their SAT scores and geographic location, but also their personalities.

There are a few possible directions for future work. First, we can investigate a fine-grained partition strategy to divide the data by departments, subjects or courses. Second, we can track the comments over time. Our current dataset contain comments from 2018 to 2020, with most of the comments posted in May and December, which are the ends of spring and fall semesters. With more data over time, we may study individual professor's teaching style changes and look at

this problem from a temporal point of view. Finally, many in-person lectures are switched to online lectures due to COVID-19 and quarantine. A valuable study would be to first determine the courses that are transformed from in-person to online and then understand the changes from the student's experiences.

References

1. Rate my professors about page (2020), https://www.ratemyprofessors.com/About.jsp
2. Abulaish, M., Jahiruddin, Doja, M.N., Ahmad, T.: Feature and opinion mining for customer review summarization. In: Chaudhury, S., Mitra, S., Murthy, C.A., Sastry, P.S., Pal, S.K. (eds.) Pattern Recognition and Machine Intelligence. pp. 219–224. Springer Berlin Heidelberg, Berlin, Heidelberg (2009)
3. Arbaugh, J.: How instructor immediacy behaviors affect student satisfaction and learning in web-based courses. Business Communication Quarterly **64**(4), 42–54 (2001)
4. Blei, D., Ng, A., Jordan, M.: Latent dirichlet allocation. Journal of Machine Learning Research **3**, 993–1022 (05 2003)
5. Blei, D.M., Lafferty, J.D.: Dynamic topic models. In: Proceedings of the 23rd International Conference on Machine Learning. p. 113–120. ICML '06, Association for Computing Machinery, New York, NY, USA (2006)
6. Chen, L., Lyu, H., Yang, T., Wang, Y., Luo, J.: In the eyes of the beholder: Analyzing social media use of neutral and controversial terms for covid-19 (2020)
7. Etzioni, O., Cafarella, M., Downey, D., Popescu, A.M., Shaked, T., Soderland, S., Weld, D.S., Yates, A.: Unsupervised named-entity extraction from the web: An experimental study. Artif. Intell. **165**(1), 91–134 (Jun 2005)
8. Hu, M., Liu, B.: Mining and summarizing customer reviews. p. 168–177. KDD '04, Association for Computing Machinery, New York, NY, USA (2004)
9. Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.J.: Introduction to WordNet: An On-line Lexical Database*. International Journal of Lexicography **3**(4), 235–244 (12 1990)
10. Monks, J., Ehrenberg, R.G.: 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)
11. Morse, R., Brooks, E.: How u.s. news calculated the 2021 best colleges rankings (Sep 2020), https://www.usnews.com/education/best-colleges/articles/how-us-news-calculated-the-rankings
12. Newman, D., Asuncion, A., Smyth, P., Welling, M.: Distributed algorithms for topic models. J. Mach. Learn. Res. **10**, 1801–1828 (Dec 2009)
13. Otto, J., Jr, D.A.S., Ross, D.N.: Does ratemyprofessor.com really rate my professor? Assessment & Evaluation in Higher Education **33**(4), 355–368 (2008)
14. Otto, J., Sanford, D., Wagner, W.: Analysis of online student ratings of university faculty. Journal of College Teaching & Learning (TLC) **2**, 25–30 (01 2011)
15. Pennebaker, J., Boyd, R., Jordan, K., Blackburn, K.: The development and psychometric properties of liwc2015. University of Texas Libraries (2015), http://hdl.handle.net/2152/31333
16. Silva, K.M., Silva, F.J., Quinn, M.A., Draper, J.N., Cover, K.R., Munoff, A.A.: Rate my professor: Online evaluations of psychology instructors. Teaching of Psychology **35**(2), 71–80 (2008)
17. Titov, I., McDonald, R.T.: Modeling online reviews with multi-grain topic models. CoRR **abs/0801.1063** (2008), http://arxiv.org/abs/0801.1063