Ashley Miller, Jason Crismore, Nathaniel Moody

Project Proposal

25 September 2017

Introduction

While a multitude of factors contribute to college selection among students, national and even international ranking lists can be used as a technique to shorten the initial consideration set (Hazelkorn 2014). Many list types and methodologies exist; however, the *U.S. News and World Report (USNWR)* rankings popularized in the 1980s started the process of looking beyond the research reputation of universities. Instead, this ranking system takes into consideration graduation and retention rates, faculty resources, alumni giving rates, as well as undergraduate academic reputation through a survey assessment that currently accounts for 22.5% of the ranking calculation (Morse, 2017). Academic reputation of a university can be conveyed in a number of ways via traditional media methods as well as through social media. Today, one of out every five high school students utilize Twitter to learn more about a particular college (Geyer, 2017*)*. Colleges and universities have now included social media as an integral part of their university's 'brand' strategy which calls into question, how much of a role does social media play into the rankings game (Rutter, 2016).

As of July 2017, there are now nearly "2.4 million fewer college bound students" which means increased competition among higher education institutions (Carey, 2017). Social media has now been cited as a "phenomenon which can drastically impact a brand's reputation and in some cases survival" (Rutter, 2016). Prior research reviews show that while social media can impact the performance of a university in regards to branding, affinity, and recruitment, little data is available to showcase the relationship, if any, between social media and national rankings. Since timing of rankings are released annually, this project methodology could be used as a way to continually track performance over time in between ranking releases so institutions can better gauge where they may fall in the mix of other colleges and universities. Further, this analysis may provide guidance on their overall marketing and branding efforts. With these factors in mind, this project intends to explore whether the use of Twitter data can be used as a method to predict a college or university's USNWR ranking, which will be used as a proxy for 'brand strength'.

Literature Review

Several other studies and sources have broached the topic of what relationship, if any, exists between social media as it relates to university branding and college rankings. The use of available Twitter data has been used in exploring the relationships among top-ranked universities to determine whether universities with common features are likely to associate with each other (Shields, 2016). While this study showed a relationship between institution rankings to other 'world-class' universities, it also stated there was a relatively small impact. Further, this study only evaluated accounts of the top-ranked universities themselves and did not explore perception or sentiment among those in the 'general population'. While the use of data points such as followers, hashtags, and mentions used in this study will likely provide guidance into our work, we hope to fill the void left by this study in examining similar data among a more representative audience. Existing research has also explored whether social media use can be tied to student outcomes and performance. One experimental study showed the positive impact of Twitter use on academic performance and engagement. Specifically, this study did not address the population of students included so it begs the question whether one is truly influenced by the other or whether these students were already 'high-achievers', with or without the use of Twitter (Davis, 2012).

Other work has showcased the ability to recreate the ranking methodology and adjust factors within to determine which variable(s) would need to be achieved in order to reach a specific ranking. An issue of 'noise' is addressed frequently as a reason why rankings would fluctuate over time (Gnolek, 2014). Evaluating sentiment data on Twitter may provide one opportunity to address this issue of 'noise' in the data that is not addressed in great detail from other research. While Hazelkorn states that the influence of rankings on public consciousness is 'immeasurable', one could argue that the use of social media could be used in an effort to fill this gap and provide a consistent and measurable process to utilize over time (Hazelkorn, 2014).

Research by Rutter, Rope, and Lettice discuss a similar question to ours as they evaluated the relationship of higher education institution's use of social media and inclusion in the Russell Group study (Rutter, 2016). While this study found that "universities that interact more with their followers achieve better student recruitment performance than universities that fail to interact", this merely looked at the number of interactions versus the quality or sentiment of interactions to further validate the relationship (Rutter, 2016). Clark, Fine, and Scheuer sought to understand the relationship current students had with their university through not only social media data but also with primary quantitative research (Clark, 2017). This approach seeks to explore the sentiment around the university and relationship quality – which other work does not address in

detail. However, sample size utilized was below the threshold of obtaining a five percent margin-of-error and convenience-based. Further, the authors only targeted those in marketing courses at the university which does not allow the ability to be representative of the student population and therefore results may be skewed. Also, their focus was only on one institution which limits the ability to determine trends. Our approach intends to explore multiple institutions to better understand the sentiment of social media as it relates to rankings and what information can be gleaned from that relationship.

Dataset Description

The ideal dataset for this project will be drawn from Twitter status updates across the United States that pertain to those universities ranked by US News. Specifically, this project will focus on the rankings received by schools in the Big Ten conference. This data is readily available, and easy to access. To acquire this data, the Twitter API will be utilized. Status updates will be collected over the course of one week, at different points during the day, and searched based on the names of Big Ten universities from the US News National University Rankings list ("2018 Best National Universities", 2017).

Once collected, this dataset should form a corpus of Twitter status updates that reference the ranked universities from a representative sample of Twitter users. This dataset will control for possible confounding variables such as time-of-day, day-of-the-week, or excessive attention to a daily current-event (i.e. a football game) by spreading out the collection process and gathering data from across the country.

After collection, this dataset will be cleaned and features will be selected from it, resulting in data that is ready for analysis. Features selected from the data will include date/time, text, and location. The text from each status, specifically, will be evaluated and cleaned to ensure its readiness for natural language processing and sentiment analysis. The date/time and location features will be preserved alongside the text for possible consideration in the discussion of this project's results.

Research Design and Methods

The general thrust of this project is to test the accuracy of social media sentiment in predicting the rankings that are received by major universities. To that end, simple natural language processing and sentiment analysis will be used to rank the representative polarity of social media activity toward the ranked universities.

The text from the dataset described above will be transformed and analyzed, particularly. Each Tweet's text feature will be broken down ("tokenized") and natural language processing will be

used to rank the polarity of each word, by employing a lexicon, such as the SentiWordNet lexicon found at http://sentiwordnet.isti.cnr.it/ (SentiWordNet, 2017). The overall sentiment of the Tweet will then be calculated using these polarity scores, and the overall sentiment of social media activity toward each ranked university will be individually summarized from these collection of Tweet sentiments across the entire dataset. Stated more simply: The sentiment of each Twitter status toward the university being referenced will be quantified and added up.

Once this sentiment analysis is performed, a model will be created, a so-called "Social Media Ranking" of the universities being considered formed by comparing their polarity scores, ranked from most-positive to least. This ranking will then be compared to the rankings generated by US News, as a measure of the model's accuracy. Model accuracy will be considered based on how much the "Social Media Ranking" differs from the US News rankings. Generally, the closer our model is to the rankings created by US News, the better the model's performance will be considered.

Conclusions

This project proposal has implications for the outreach and recruiting strategy at America's universities. If the recruitment at a university could be predicted or artificially manipulated by social media interactions, then colleges aware of causal relationship could take advantage of it and drive recruitment numbers upwards. This would disadvantage competing institutions.

The model chosen for this proposal is designed to provide a reasonable prediction of the correct ranking of universities based on Twitter usage statistics. This data is readily accessible and can be managed within the timeframe of this course. The proposal also identified a reasonable benchmark published by U.S. News & World Report that already provides rankings for the top colleges and universities in the United States. This ability to show past research, conduct current research, and compare against a national standard provides a framework that should contribute to the body of knowledge in social mining research.

References

2018 Best National Universities (2017). Retrieved September 24, 2017, from https://www.usnews.com/best-colleges/rankings/national-universities

Carey, I., & Marcus, J. (2017). There are 2.4 million fewer college students than there were five years ago - The Hechinger Report. Retrieved September 24, 2017, from http://hechingerreport.org/2-4-million-fewer-college-students-five-years-ago/

Clark, M., Fine, M. B., & Scheuer, C.-L. (2017). Relationship quality in higher education marketing: the role of social media engagement. Journal of Marketing for Higher Education, 27(1), 40–58. https://doi.org/10.1080/08841241.2016.1269036

Davis III, C. H. F., Deil-Amen, R., Rios-Aguilar, C., & Canche, M. S. G. (2012). Social Media in Higher Education: A literature review and research directions. The Center for the Study of Higher Education at The University of Arizona, 23.

Geyer, S. (n.d.). Ruffalo Noel Levitz E-Expectations 2017: How Online Expectations Change for Today's Prospective Students and Parents. Retrieved from http://www.act.org/content/dam/act/unsecured/documents/epc2017/Session-T4.5-E-Expectations-2017_v2.pdf

Gnolek, S. L., Falciano, V. T., & Kuncl, R. W. (2014). Modeling Change and Variation in U.S. News & World Report College Rankings: What would it really take to be in the Top 20? Research in Higher Education, 55(8), 761–779. https://doi.org/10.1007/s11162-014-9336-9

Hazelkorn, E. (2014). Rankings and the Global Reputation Race. New Directions for Higher Education, 2014(168), 13–26. https://doi.org/10.1002/he.20110

Morse, R., Brooks, E., & Mason, M. (2017). How U.S. News Calculated the 2018 Best Colleges Rankings. Retrieved September 24, 2017, from https://www.usnews.com/education/best-colleges/articles/how-us-news-calculated-the-rankings

Rutter, R., Roper, S., & Lettice, F. (2016). Social media interaction, the university brand and recruitment performance. Journal of Business Research, 69(8), 3096–3104. https://doi.org/10.1016/j.jbusres.2016.01.025

SentiWordNet. (2017). Retrieved September 24, 2017, from http://sentiwordnet.isti.cnr.it/

Shields, R. (2016). Following the leader? Network models of "world-class" universities on Twitter. Higher Education, 71(2), 253–268. https://doi.org/10.1007/s10734-015-9900-z