Yelp Review Manipulation and Social Effects

Elaine Arbaugh
California Institute of Technology
Pasadena, CA
earbaugh@caltech.edu

Zachary Fein
California Institute of Technology
Pasadena,CA
zfein@caltech.edu

ABSTRACT

Yelp has been put under fire recently, as small business owners claim that Yelp pressures them into advertising in exchange for more positive reviews; if they refuse to advertise, they notice more negative reviews on their page. Despite these accusations, Yelp insists that advertisers do not get higher ratings, their negative reviews removed, or more recommended positive reviews. For this project, we examine Yelp restaurant review data from 578 restaurants in downtown Los Angeles to try to find evidence of review manipulation. Information for more than 130,000 reviews was collected. Using logistic regression, we find no evidence that advertisers have more higher-than average reviews at the top of their page than non-advertisers. Hoping to find causal evidence instead, we found four restaurants who started advertising on Yelp during the course of our project, but there were no obvious signs that their reviews had been manipulated after they began advertising. With our wealth of data, we also focus on identifying information cascades in Yelp reviews, which occurs when a user rates a restaurant based on the reviews he sees rather than his own experience. We find signs of cascading behavior in ratings, which we argue is evidence of social and cascade effects rather than exogenous factors. In most of our analyses, data on a more long term scale is required to truly see how reviews and ratings change over time.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; H.2.8 [**Database Applications**]: Data Mining
; G.3 [**Probability and Statistics**]: Correlation and regression analysis

Keywords

Yelp, Data mining, Information cascades

1. BACKGROUND

Yelp has become extremely popular after its creation in 2004. In 2014, it had 135 million monthly unique visitors, as well as 71 million total page views. According to a 2014 Nielsen survey, 44% of consumers user Yelp to search for local businesses. 90% of consumers say their choices are influenced by positive Yelp reviews, while 86% are influenced by negative reviews, according to a 2013 Dimensional Research study [6]. So, Yelp reviews have a large influence on consumers' choices, and the reviews also influence a company's popularity and profit. In Michael Luca's study "Reviews, Reputation, and Revenue: The Case of Yelp.com," he found that for each additional star rating a business had on Yelp, revenue increased by between 5 and 9 percent [3].

1.1 Review Manipulation Claims

There has been a large amount of controversy surrounding Yelp reviews. Many small business owners have claimed that Yelp manipulates reviews so that non-advertising businesses have negative reviews in prominent locations, while advertising companies have favorable reviews in prominent locations. For instance, one restaurant owner in the Bay Area claimed he received a call from a Yelp salesman who said that the restaurant's positive reviews could be moved to the top of the page if he paid $299 a month for advertising [7]. Other small business owners have claimed to have received similar calls from Yelp, demanding payment for advertising in exchange for hiding negative reviews and promoting positive reviews. In addition, some business owners have claimed that after they refused Yelp's offer, they received more negative reviews and fewer positive ones [7]. Yelp received over 2000 Federal Trade Commission complaints between 2008 and 2014, so these cases are not isolated [2].

Despite what these small business owners say, Yelp claims on their website that "Money doesn't buy anything but ads," and that advertisers do not get higher ratings, are not allowed to remove negative reviews, and do not get an increased number of recommended positive reviews [9]. All lawsuits against Yelp related to review manipulation have been dismissed.

1.2 Yelp Page Overview

Yelp sorts its reviews into recommended and filtered reviews, where filtered reviews are hidden and do not factor into the business's rating or review count. Filtered reviews are only accessible by clicking on a difficult to notice link at the bottom of a business's Yelp page. Up to 25% of all reviews on the site are not recommended, and some businesses have the majority of their reviews filtered [3]. Yelp's website says that reviews are sorted by an automatic filter which takes into account "various measures of quality, reliability, and activity," and that the reviews they want to filter are fake, posted by non-established reviewers, or unhelpful [8]. Yelp admits that they filter many legitimate reviews, which they justify because it is preferable to having fake reviews on the site. Details about the review filtering algorithm have not been released to the public.

Each company's page shows all of its reviews, sorted by default by Yelp Sort, Yelp's page rank algorithm. According to their website, the ordering is determined by "recency, user voting, and other review quality factors" [11]. The first review on a page is supposed to be "one that reflects the average star rating of a business" [10]. Yelp's review rank algorithm is also not publicly available.

The primary complaints about Yelp review manipulation were that Yelp Sort promoted positive reviews for advertising companies and negative reviews for companies who did not advertise, and that Yelp's filtering algorithm filtered negative reviews for advertising companies and positive reviews for non-advertising companies.

1.3 Information Cascades

Information cascades occur when people act based on what other people do, rather than their own private information. In the case of Yelp reviews, a cascade would entail people basing their ratings on top reviews, rather than what they actually think of the business. For instance, if a customer has a terrible experience at a restaurant and plans to give it a 1 star rating, but sees that all of the top five reviews are 5 star ratings, he may give the restaurant a higher review because he believes his experience was an aberration. If many reviewers in a row engage in this behavior, a business may receive an artificially inflated or deflated overall rating.

1.4 Goals

We planned to gather and analyze data to find out whether Yelp manipulates reviews by ranking positive reviews higher or negative reviews lower for advertisers, as well as to see if there was any evidence of information cascading in the reviews.

2. RELATED WORKS

Several researchers have previously analyzed the Yelp review filter and categorized types of reviews which often get filtered.

2.1 "Fake It Till You Make It"

In the paper "Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud," Michael Luca and Georgios Zervas examined the factors that caused reviews to be filtered. Importantly, they found that Yelp does not filter reviews differently for advertisers compared to non-advertisers. They trained a linear regression to predict whether a review would be filtered based on the reviewer's characteristics as well as the star rating of the review and the advertising status of the restaurant; the advertising terms were not significant, so they conclude no signs of filtering manipulation for advertisers [4]. We will use a similar approach later in this paper to examine claims of Yelp Sort being biased towards advertisers.

2.2 "Understanding the Yelp Review Filter"

In the 2014 paper "Understanding the Yelp review filter: An exploratory study," David Kamerer used data from restaurants and religious organizations in Chicago to analyze Yelp's filtering process. Kamerer compared characteristics of reviews, like word count and date, and characteristics of reviewers, like number of reviews and number of friends, for filtered and visible reviews. Kamerer found that factors related to the reviewer are strongly correlated with filtering, but that factors related to the review are not [1].

2.3 "What Yelp Fake Review Filter Might Be Doing?"

In the paper "What Yelp Fake Review Filter Might Be Doing?," Arjun Mukherjee, Vivek Venkataraman, Bing Liu, and Natalie Glance analyzed Yelp's filtered reviews to model Yelp's filtering algorithm. They created a supervised learning model using different feature sets (LIWC + bigrams, style and part-of-speech, and deep syntax) to classify filtered reviews. Their model ended up having 67.8% accuracy [5].

3. DATA COLLECTION

We gathered restaurant and review data from 578 restaurants in the downtown Los Angeles area, including 71 advertising restaurants. The data included 134,564 recommended reviews and 14,361 filtered reviews. Our plan was to find differences between the reviews advertising and non-advertising restaurants received, as well as to find restaurants who became advertisers after we started data collection and analyze how their review rankings and filtering changed.

3.1 Methodology

We could not use the Yelp API, since it provides no information about whether or not a company is an advertiser and limited information about reviews. Instead, we scraped pages using Python with BeautifulSoup and urllib2 for the top restaurants in downtown Los Angeles. We collected almost all available information about each restaurant and review, even if we thought the data would be irrelevant, and stored it in a MySQL database instance on AWS. The data we collected for each restaurant and review is described in Tables 1 and 2.

We collected an initial round of data on April 21, 2015 for every restaurant in our data set. We also gathered data for restaurants which were not advertisers on April 21 but became advertisers four times: May 5, May 20, May 26, and June 10.

3.2 Issues

We encountered several problems when scraping pages. We initially set our crawler to have a 30 second pause time in between page accesses for politeness. However, this caused our scripts to be extremely slow, since each business has 40 reviews per page, and some businesses have up to 10,000 reviews. With the 30 second wait time, collecting data for a

Table 1: Restaurant Characteristics

Name	Address
Food type	Number of Photos
Dollar Sign Rating	Advertiser?
Star Rating	Number of Reviews
Takes Reservation?	Delivery
Take-out?	Accepts Credit Card?
Accepts Bitcoin?	Good for?
Parking	Bike Parking
Wheelchair Accessible	Good for Kids?
Good for Groups?	Attire
Ambiance	Noise Level
Alcohol	Outdoor Seating?
Wi-Fi?	Has TV?
Dogs Allowed?	Waiter Service
Drive Thru?	Caters?
Latitute	Longitude

Table 2: Review Characteristics

Review Rank	Restaurant
Name	Location
Default Avatar?	Number of Friends
Number of Reviews	Elite?
Number of Pictures	Score
Date	Useful Ratings
Funny Ratings	Cool Ratings
Checked In?	Review Text
Word Count	Average Word Length
Percent Capitalization	

business with 10,000 reviews would take over 2 hours. In addition, every couple hours Yelp would ask for a CAPTCHA to "check that you're not a robot." We experimented with reducing the wait times, and found that Yelp asked for a CAPTCHA approximately every 2 hours no matter how many times pages were accessed. We ended up using wait times of 2-5 seconds, and checking yelp.com every so often to see if a CAPTCHA was required. Yelp also changed its HTML a couple times, but it was fairly simple to update our scraping code to reflect the changes.

4. REVIEW MANIPULATION ANALYSIS

One of the main ways to test these claims of review manipulation is to determine if Yelp favors its advertisers in terms of ranking more higher-than-average reviews at the top of the page compared to non-advertisers. We cannot assume that advertisers represent a random sample of all restaurants, so we need to calculate more than just summary statistics in order to study this. For this, we used a regression to find the effects of advertising on review placement. We decided to try to predict whether a review would place in the top 10 reviews for a restaurant. We did not want to use the overall rank of a review as the dependent variable due to differing numbers of total reviews for restaurants; a 20^{th}-ranked review for a high-end restaurant with 1,000 reviews is not equivalent to the 20^{th}-ranked review for a food cart with 20 reviews. Furthermore, we did not want to use the rank percentile of a review —that is, the review's rank divided by the total number of reviews —because this method would hold for instance that reviews 951 to 1000 for that same high-end restaurant are all worse in quality than every review for that food cart except its worst-rated review. We did not feel that this assumption was valid.

4.1 Data Cleaning

Before running any regressions, we first cleaned up the data we had to use as features. For each review, we combined the useful, funny, and cool feedback ratings into one number, since they all signified that other Yelp reviewers thought the review was beneficial. Since we were interested in how reviews' ratings compared to the rest of the reviews for the same restaurant, rather than their absolute ratings, we created a binary indicator for whether the user's rating was greater than the restaurant's average rating.

4.2 Logistic Regression

We wanted to make a model to predict the binary outcome of whether a particular review would be ranked in the top 10 reviews for a restaurant. Our response variable is binary —1 if the review is ranked in the top 10 and 0 otherwise —so standard linear regression is not applicable. We used logistic regression instead, which naturally bounds the predicted values between 0 and 1 and so essentially predicts the probability of the response variable occurring. Regression models serve the purpose of returning the marginal effect of each feature on the output variable, so this allows us to determine how much advertising plays a role in whether a review will be highly-ranked.

Even though we had collected over 40 features for each review, not all of them would be useful in our logistic regression model. A preliminary logistic regression was run using all the features gathered as explanatory variables, and the most irrelevant features, as determined by the p-value of the coefficients corresponding to each feature, were discarded. This left us with a large list of features, all of which had some effect on review placement.

4.2.1 Feature Selection

We selected the features for the final model as follows. First, we removed the advertising indicator in order to find the best features disregarding advertising status. Then, given every combination of the remaining features, we wanted to find the set that gave us the best model fit to use in the final model. We define "best model fit" using F1 score, which is a metric that rewards for a balance in the rate of true positives, false positives, true negatives, and false positives. (In our problem, a true positive would be an example of a review that actually placed in the top 10 and was also predicted to do so by the regression and a false positive would be an example of a review predicted to be in the top 10 but was actually placed outside, and similar examples hold for the other two cases.) The logistic regression method actually tries to minimize the root-mean-square-error (RMSE) between the predicted probability and the actual 0/1 outcome; however, since there were so few reviews that ranked in the top 10, a model could output 0 for every review and possibly have a better RMSE than a normal model. Thus, we decided to choose the set of features that minimized F1 score to use in the final model.

Table 3: Logistic regression to predict review's rank $\leq$ 10

	(1)		(2)	
Constant	-82.99***	(1.536)	-83.43***	(1.539)
User is elite?	1.299***	(0.039)	1.295***	(0.039)
User has default avatar?	-1.240***	(0.081)	-1.239***	(0.081)
Number photos in review	0.3634***	(0.017)	0.367***	(0.017)
Date epoch	$5.78 \text{x} 10^{-8}$***	($1.09 \text{x} 10^{-9}$)	$5.82 \text{x} 10^{-8}$***	($1.09 \text{x} 10^{-9}$)
Number restaurant reviews	-0.0032***	($6.78 \text{x} 10^{-5}$)	-0.0032***	($6.71 \text{x} 10^{-5}$)
Useful + funny + cool feedback ratings	0.0087***	(0.001)	0.0084***	(0.001)
User rating >restaurant rating?	-0.133***	(0.035)	-0.133***	(0.038)
Restaurant is advertiser?			-0.535***	(0.073)
(User rating >restaurant rating?) x (advertiser?)			0.0595	(0.094)
N	126943		126943	
R^2	0.4017		0.4049	
*** p <0.001				

4.2.2 Validation

We used 10-fold cross validation for this task so that we could find the combination of features that best generalizes to out-of-sample data rather than those that best fit on the data set we collected. We split our data into 10 equally-sized sections or "folds". Then, for each combination of features, we trained a logistic regression model using nine of the 10 folds as the explanatory data and the other fold as the response data and then calculated the F1 score between the predicted values for the response data and the actual data; we repeated this 10 times so that each fold was used as the response values exactly once. Then, the average F1 score over the 10 iterations was calculated. The set of features with the highest average F1 score was determined to be the best-fitting set of features.

4.2.3 Advertising Effects

To measure advertising effects, we used two added features: the indicator variable of whether the restaurant being reviewed advertised on Yelp as well as the interaction term between this indicator and the indicator of whether the review's score was above average. The interaction term will tell us whether advertisers are given benefits with regards to more or fewer higher-than-average ratings at the top of their page: If the coefficient for this variable is significantly above 0, then that suggests that higher-than-average reviews for advertising restaurants have a statistically significantly better chance of being ranked at the top of the page than either lower-than-average reviews or similarly-rated reviews for non-advertisers. We add in the advertiser indicator also so that any advertiser effects independent of rating (such as, for example, advertising restaurants having more reviews and so a lower chance for their average review to be in the top 10) are taken into consideration separately and are not wrapped into the interaction term.

We chose this approach instead of running two separate regressions, one with advertising restaurants and one with non-advertising restaurants, and comparing the resulting coefficients. The two-step method is essentially the same as running one regression and including interaction terms between advertising status and each of the original features in the model. (Any differences in the coefficients between the two separate regressions would be used as the values of the interaction coefficients in the single regression.) However, we were not interested in whether, for instance, reviews with photos includes are more likely to be highly-ranked on advertising restaurants than non-advertising restaurants. Since accusations against Yelp have pointed to manipulation with regard to reviews' ratings, we only wanted to consider the interaction between advertising status and the review's rating above or below average.

4.2.4 Results

With the best set of features chosen, we ran the logistic regression on the entire data set with and without these advertising effects. The coefficients and standard errors for each model are shown in Table 3 below.

The advertising interaction term is not significant, and its inclusion in the model does not affect the coefficients of the other terms. In other words, while it does matter if a review's rating is higher than average (User rating > restaurant rating? is significant at $p = 0.001$ in the original model), it does not also matter if the restaurant is an advertiser. Thus, the logistic regression results show that there is no evidence that being a Yelp advertiser affects whether higher-than-average reviews will appear at the top of the restaurant's page.

5. NEW ADVERTISING RESTAURANTS

We wanted to find restaurants who started advertising after we started collecting data to see if their review rankings of filtered reviews changed after they became advertisers. We found four restaurants who started advertising after April 21: Poppy + Rose, El Huitlacoche Restaurant Mexican Food, Preux & Proper, and Palm Restaurant. Poppy + Rose became an advertiser between April 21 and May 7, while El Huitlacoche became an advertiser between May 7 and May 20 and both Preux & Proper and Palm Restaurant became advertisers between May 26 and June 10.

5.1 Poppy + Rose

A summary of Poppy + Rose's reviews before and after they started advertising is in Table 3. A running average of the top x reviews for different dates is given in Figure 1.

Table 4: Poppy + Rose New Filtered Reviews

Name	Rating	Date	# Reviews	# Friends	PR[1]	Review Text
Josh R.	4	4/22/2015	15	0	No	Good food but delivery took a long time and food was cold-ish
Simon B.	4	6/2/2015	5	35	No	It's an oasis. They got a real nice batter for their fried chicken sandwich, good coffee by Frontier and they serve Brooklyn Bagels.
Albert B.	4	4/22/2015	5	1	Yes	My office friend took me here and i must say that food here is delicious. We ordered fried chicken here and i must say it was well cooked and fresh. The portions are a good size. Nice and clean place though it was super packed. Will definitely be going back to try some other menu items.
Rodrigo B.	5	4/26/2015	1	0	No	My wife and I were in the flower market and decided to eat in the restaurant and were amazed by how great their breakfast was! The breakfast sandwich is amazing! Their almond croissant is perfectly flaky and full of flavor! Lastly their two egg special is fantastic! Best hash browns I've ever eaten!

Table 5: Poppy + Rose Statistics

	4/21	5/7	5/20	5/26	6/10
Advertiser	No	Yes	Yes	Yes	Yes
Review Count	216	232	251	258	279
Average Rating	4.26	4.25	4.24	4.19	4.19
Top 10 Average	4.2	4.3	3.9	3.9	4.1
Top 5 Average	3.8	4.6	3.4	3.6	4.4
Filtered Count	7	8	8	8	10
Filtered Average	3.86	3.88	3.88	3.88	3.90

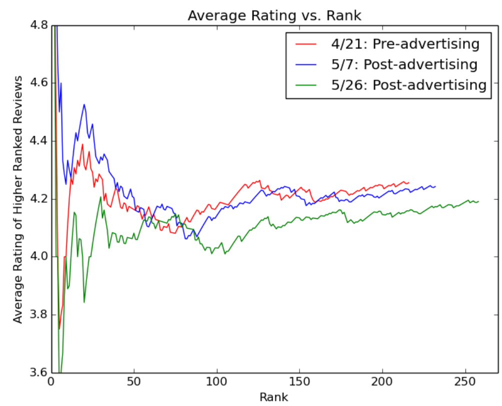

Figure 1: Review average of the top x reviews for different dates. For instance, Rank=50 would be the average of the first 50 reviews.

5.1.1 *Poppy + Rose Top Reviews*

The average rating for top reviews initially increased after Poppy + Rose started advertising, from 4.2 (top 10) and 3.8 (top 5) on April 21 to 4.3 (top 10) and 4.6 (top 5) on May 7. So, it initially seemed like advertising may have improved Poppy + Rose's top reviews. However, both averages then decreased on May 20 and May 26 before increasing again on 6/10. So, due to daily variation in review ordering, it is hard to tell whether or not becoming an advertiser improved Poppy + Rose's top reviews.

5.1.2 *Poppy + Rose Filtered Reviews*

Poppy + Rose received 4 new filtered reviews after they started advertising, while one review, written by Josh M., was moved from the filtered section to the recommended section. The filtered reviews are described in Table 4.

All of the reviews filtered after Poppy + Rose started advertising seem to have been filtered legitimately. They all give the restaurant 4 or more stars, supporting the idea that Yelp does not filter negative reviews for their advertisers. In addition, all are fairly short, and several have grammatical errors (lack of punctuation in Josh R.'s, "They got" in Simon B.'s, and lowercase "i" in Albert B.'s). None of the reviewers have written very many reviews, and all except Simon B. have few friends. So, based on the characteristics of the reviews and the reviewers, Yelp filtering these reviews for Poppy + Rose is reasonable.

Josh M.'s review, which was moved from the filtered section to the recommended section, is about the same caliber as the other filtered reviews–it says "The chicken is amazing. Crispy and tasty batter. A bit of a wait because this place is quite popular. I had the Lumberjack breakfast. Can't go wrong if you're a traditional breakfast lover. The service is really good. I am definitely coming back." Although Josh M. did write more reviews between April 21 and June 10, he only wrote 5 total, a similar amount as the users who wrote the other filtered reviews. He had 21 friends on June 10, which is less than Simon B., whose review was filtered. So, it is unclear why Josh M.'s review was unfiltered while some of the others were not.

Table 6: Preux & Proper New Filtered Reviews

Name	Rating	Date	# Reviews	# Friends	Previously Recommended?
Foie G.	4	5/2/2015	5	0	No
Don Juan S.	5	4/28/2015	29	50	No
Lee E.	5	4/1/2015	4	0	Yes
Shaun H.	5	3/31/2015	1	0	Yes

Table 7: Preux & Proper Recently Unfiltered Reviews

Name	Rating	Date	# Reviews	# Friends
Stanley S.	5	4/13/2015	5	5
Jennifer P.	5	4/12/2015	2	0
Jeremy S.	5	3/23/2015	1	0
Koukla H.	5	3/22/2015	1	0

5.2 El Huitlacoche

El Huitlacoche was less interesting than Poppy + Rose, since it only had 5 recommended reviews and 2 filtered reviews when we first collected data on April 21. After it became an advertiser, it received one 5 star recommended review. Since there is so little data for El Huitlacoche, it does not tell us very much about whether or not Yelp manipulates reviews.

5.3 Preux & Proper

Data for Preux & Proper reviews before and after they started advertising is in Table 10.

5.3.1 Preux & Proper Top Reviews

The average review score for the top 10 and top 5 were both higher on June 10 than April 21, increasing from 3.3 to 3.8 and 3.4 to 4.0, respectively. The average rating also increased slightly, from 3.86 to 3.90, and the average rating for new reviews between April 21 and June 10 was 4.06. So, although the average for top reviews increased, it ended up at a level consistent with the overall average score for the time period. The average of top reviews before Preux & Proper started advertising is lower than expected, but that could be attributed to other review factors. So, there is no clear evidence of review manipulation for Preux & Proper's top reviews.

5.3.2 Preux & Proper Filtered Reviews

Although Preux & Proper had 13 filtered reviews on both April 21 and June 10, four of the reviews for each date are different–four filtered reviews from April 21 became recommended, and four new reviews were filtered. Data for the new filtered reviews is in Table 6, while data for the recently unfiltered reviews is in Table 7.

Two of the newly filtered reviews (Lee E. and Shaun H.) were previously recommended on 4/21. In addition, one review by Andy F., which said "Don't go. Fried food cover up. Would not recommend to worst enemy. Unless you are dying (literally) of hunger, don't go. Don't go. Don't go." was recommended on 4/21, but disappeared on 6/10–it was neither recommended nor filtered.

All the reviews that Yelp unfiltered between 4/21 and 6/10 were 5 stars, which is somewhat suspicious, especially since none of the reviewers were particularly active on the site in terms of friends or reviews. Andy F.'s disappearing review is even more suspicious, although it may have been deleted for a multitude of reasons: for instance, Andy could have deleted his account or violated the terms of service. Yelp also filtered two 5 star reviews which were previously recommended, so there is no clear trend of Yelp filtering reviews with low ratings and recommending reviews with high ratings.

Table 10: Preux & Proper Statistics

	4/21	6/10
Advertiser	No	Yes
Review Count	155	203
Average Rating	3.86	3.90
Top 10 Average	3.3	3.8
Top 5 Average	3.4	4.0
Filtered Count	13	13
Filtered Average	4.38	4.31

Table 11: Palm Restaurant Statistics

	4/21	6/10
Advertiser	No	Yes
Review Count	406	407
Average Rating	3.56	3.56
Top 10 Average	4.2	3.9
Top 5 Average	3.8	3.6
Filtered Count	54	59
Filtered Average	3.59	3.54

5.4 Palm Restaurant

Table 11 shows review data for Palm Restaurant before and after they began advertising.

5.4.1 Palm Restaurant Top Reviews

Both the top 10 and top 5 average decreased after Palm Restaurant after they began advertising, although the average rating stayed approximately the same.

5.4.2 Palm Restaurant Filtered Reviews

Filtered review data for Palm Restaurant is in Tables 8 and 9. Between April 21 and June 10, Yelp recommended one

Table 8: Palm Restaurant New Filtered Reviews

Name	Rating	Date	# Reviews	# Friends	Previously Recommended?
Matt G.	2	12/4/2010	9	0	Yes
Gino I.	3	5/22/2015	153	322	No
Yelps S.	3	4/26/2015	109	1	No
Eddie P.	4	6/10/2015	3	17	No
Amanda H.	4	6/2/2013	23	0	Yes
Danielle T.	4	1/16/2011	15	0	Yes

Table 9: Palm Restaurant Recently Unfiltered Reviews

Name	Rating	Date	# Reviews	# Friends	Review Text
Evelina K.	5	5/8/2013	21	0	One of the best steaks in LA. Should go there hungry. Its love in every bite!

previously filtered 5-star review for Palm Restaurant, while filtering one 2 star review, two 3 star reviews, and three 4 star reviews. Notably, the 2 star review and two of the 4 star reviews were previously recommended. These three reviews are also fairly old, as they were written in 2010, 2011, and 2013, so it is odd that they are being filtered now.

Also, two of the filtered reviews were written by very active Yelpers, Matt G. and Yelps S., who have both written over 100 reviews. Their reviews, although poorly written, did not look like spam, so it seems strange that reviewers who have written so many reviews are being filtered.

Palm Restaurant's filtered review activity is a bit suspicious because lower rated reviews that were written by very active Yelpers or were previously recommended are being filtered, while a short 5 star review by a less active reviewer was moved from the filtered section to the recommended section.

5.5 New Advertisers Conclusions

Although some of the data we found for the new advertising restaurants was suspicious, including increases in top 5 and top 10 star ratings for Poppy + Rose and Preux & Proper and filtering negative reviews for Palm Restaurant, the data we have is not sufficient to make any conclusions about whether or not Yelp manipulates reviews. There is no clear trend in the average rating for top reviews, since the averaged increased for Preux & Proper, decreased for Palm Restaurant, and increased, then decreased, then increased for Poppy + Rose. There is also no clear trend for which reviews were filtered and which were recommended.

One thing that surprised us was that Yelp filtered previously recommended reviews and recommended previously filtered reviews. Recommending previously filtered reviews makes sense, since Yelp filters reviewers they believe are not credible, and increased activity on the site throughout time makes the reviewer seem more legitimate. However, in the examples of Yelp recommending previously filtered reviews, we did not actually see very much evidence that the users increased their activity substantially. Filtering previously recommended reviews seems more suspect, especially when the reviews are from up to 5 years ago. Although reviews from 5 years ago may be outdated, Yelp leaves them up in most cases, so it is strange that a review which has been recommended for 5 years would suddenly be filtered.

To get more conclusive results about what happens when a restaurant becomes a Yelp advertiser, we would need to collect more data over time. First, we would have to collect data from many more restaurants–out of the about 500 non-advertising restaurants we looked at, only 4 started advertising in the month and a half we collected data for, which is not a statistically significant sample. To find a significant number of advertising restaurants, we would have to collect data for many more restaurants.

In addition, it would be helpful to collect data for all the restaurants every day or so. As we saw with Poppy + Rose, the top reviews on a page can vary greatly from day to day. So, we would want more daily data to see if reviews tended to be better after the restaurant became an advertiser. Daily data collection would also allow us to pinpoint specific events better–for instance, we know that Poppy + Rose became an advertiser between April 21 and May 7, but we do not know exactly when. If we had the time and resources, it would be interesting to try to find more general trends for newly advertising restaurants.

6. REVIEW CASCADES

With all the data we had gathered, we wanted to do more than just review manipulation analysis. We chose to examine whether an information cascade existed in Yelp reviews, in which users would rate restaurants based not only on their personal experience but also on the reviews they see before submitting their review. In this situation, a user whose restaurant experience is worthy of a 3-star review may second guess himself if he sees an influx of 5-star reviews at the top of the page and instead give the restaurant a 4-star review. If a review cascade existed, then the ratings over time for a restaurant may look like Fig. 2 —long spikes and dips as users follow the ratings of others they see.

6.1 Methodology

Unfortunately, identifying information cascades in Yelp reviews was difficult because we did not have information about the top reviews at the time each reviewer was rating the restaurant. However, we noticed that many of the top reviews were some of the most recent ones submitted. We found that top-ranked review was in the 10 most recent reviews 90% of the time as well as 92% for the #2 review and 81% for the #3 review. This tells us that when a user submits a review, the top reviews will likely be among the 10

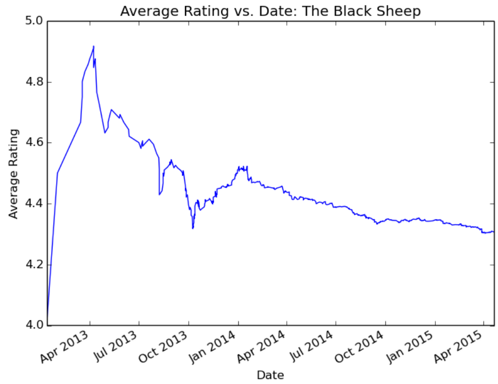

Figure 2: Example of cascading behavior in rating over time for The Black Sheep.

most recent. We will use the recency of reviews as a proxy for visibility.

6.2 Evidence for Cascades

We first looked at summary statistics to try to determine whether the cascade effect is possibly apparent in Yelp reviews. For each rating, we subtracted the average rating score up to that point to find the offset from the mean. Then we found the average offset of the 10 most recent reviews each time a new review was posted. We averaged these values over each star rating of the new review. The results are shown in Table 12 below. For instance, when a 1-star review is published, the 10 previous submitted reviews were on average 0.076 stars below the average star rating of the restaurant.

The table shows that better scores are preceded by better-than-average scores, and worse scores are preceded by worse-than-average scores. One reason for this may be due to an inherent information cascade in Yelp reviews. However, it is difficult to conclude the presence of an information cascade just from the table above because it ignores exogenous factors that may be present, such as a change in restaurant quality over time: If a 2-star restaurant revamps their menu and receives a large string of 5-star reviews as a result, this would appear as a cascade effect in Table 12.

6.3 Exogenous Factors

We analyzed review placement and recency, finding that 60.3% of the time, the most recent review was in the top 3 ranked reviews, while 59.8% of the second-most recent reviews were in the top 3 ranked reviews, and only 29.8% of the third-most recent reviews were in the top 3 ranked reviews. Based on this analysis, we believe that the cascade effect would be apparent in the previous two reviews only, while exogenous factors would emanate throughout the reviews for a long time. That is, a user has a much greater chance of seeing the two most recent reviews, so he would only change his score based on those two, whereas if, for instance, the restaurant had changed ownership and started serving higher-quality food, all of the reviews since that change would have an above-average rating, so the "cascade effect" for this scenario would be more gradual, seen through correlations between a review's score and many of the previous scores.

To study this, we wanted to determine the influence of the most recent reviews on the next review's score. We would ideally look at the visible reviews at the time of posting, but since this data was not available, we instead use recency as de facto visibility based on the numbers given above. If we find that the previous two review scores have a much larger influence on the next review's score than all of the reviews before those two, we believe that this would provide evidence of an information cascade in Yelp reviews for the reasons stated above. We determined influence through a linear regression to predict rating i given ratings $i-1$, $i-2$, ..., $i-10$, which would return coefficients that signify the influence of each of the previous 10 ratings on the next one. Each rating was mean-shifted by the restaurant's rating up to the point of the review before being included in the regression. The coefficients and standard errors of this regression are shown in Table 12.

The rating of a review depends more on the two previous ratings than the eight before them, and since the standard errors on the coefficients are so low, this difference is statistically significant. While our analysis would benefit greatly from knowing the exact reviews at the top of the page at the moment each new review is posted, the assumptions we made above are reasonable and so we have reason to believe that this is evidence of an information cascade in Yelp ratings: Users' ratings depend not only on their experience at the restaurant but also on what reviews they see before they post a new review.

Table 12: Summary statistics of cascades

Review rating	Average rating differential of the 10 previous reviews
1 stars	-0.076
2 stars	-0.042
3 stars	-0.021
4 stars	0.003
5 stars	0.025

Table 13: Linear regression of previous 10 ratings on rating i

Rating $i-1$	0.1480*** (0.003)
Rating $i-2$	0.1197*** (0.003)
Rating $i-3$	0.0927*** (0.003)
Rating $i-4$	0.1021*** (0.003)
Rating $i-5$	0.0845*** (0.003)
Rating $i-6$	0.0861*** (0.003)
Rating $i-7$	0.0852*** (0.003)
Rating $i-8$	0.0859*** (0.003)
Rating $i-9$	0.0915*** (0.003)
Rating $i-10$	0.0944*** (0.003)
N	128825
R^2	0.916
***p <0.001	

7. CONCLUSIONS

We gathered Yelp review data from over 500 restaurants in downtown Los Angeles in order to examine claims of review manipulation and to analyze potential information cascades in Yelp ratings. We did not find any convincing evidence of review manipulation for advertisers from our regressions or from looking at newly advertising restaurants. There were some changes in top review ranking and filtered reviews after the four restaurants we found became advertisers, but we do not have enough data to make any conclusions. To show causal evidence of improved ranking or filtering for advertisers, we would have to collect data over a longer period of time to find more companies who have recently become advertisers and examine how their rankings change over time. Rankings are variable from week to week, so we would need to collect data over time before and after the company starts advertising and look for general trends in positive or negative reviews. We would also have to look at a larger set of restaurants to find more advertisers, since restaurants do not become advertisers very often.

We did find evidence of cascading behavior in reviews, although we were unable to determine conclusively whether it was due to social effects or due to changes in the restaurant itself. To analyze cascading behavior further, data over time would need to be collected so we could know which reviews were the top reviews at the time instead of making assumptions about what reviews each user saw while making a new review. We could then analyze a user's behavior based on which reviews were actually the top reviews, instead of a group of reviews that they may or many not have seen. Having more data over time would also make it easier to separate exogenous effects from cascading effects.

8. ACKNOWLEDGMENTS

Thanks to Adam Wierman for all his help this term!

9. REFERENCES

[1] D. Kamerer. Understanding the yelp review filter: An exploratory study. 2014.

[2] A. Kieler. Yelp's controversial business tactics contribute to 2,000 complaints received by ftc, 2014. [Online; accessed 10-June-2015].

[3] M. Luca. Reviews, reputation, and revenue: The case of yelp.com. 2011.

[4] M. Luca and G. Zervas. Fake it till you make it: Reputation, competition, and yelp review fraud. 2015.

[5] e. a. Mukherjee, A. What yelp fake review filter might be doing? 2013.

[6] D. Research. Customer service and business results: A survey of customer service from mid-size companies. 2013.

[7] K. Richards. Yelp's controversial business tactics contribute to 2,000 complaints received by ftc, 2009. [Online; accessed 10-June-2015].

[8] Yelp. About us: Yelp, 2015. [Online; accessed 10-June-2015].

[9] Yelp. Advertiser faq: Yelp, 2015. [Online; accessed 10-June-2015].

[10] Yelp. Common questions: Yelp for business owners, 2015. [Online; accessed 10-June-2015].

[11] Yelp. How is the order of reviews determined?: Support center: Yelp, 2015. [Online; accessed 10-June-2015].