Discovering Yelp Elites: Reifying Yelp Elite Selection Criterion

Caroline Kim
University of California, San Diego
Computer Science
Email: chk101@ucsd.edu

Gordon Lin
University of California, San Diego
Computer Science
Email: golin@eng.ucsd.edu

Honam Bang
University of California, San Diego
Computer Science
Email: hobang@ucsd.edu

***Abstract*—Yelp Elite is a privileged title granted to users by Yelp. Being Elite comes with perks such as exclusive invitations to Yelp Events. Users must apply or be nominated to become Elite, and once applied, Yelp will approve and reject the application based on certain criteria. These criteria are nebulous, often leading to confused aspiring Elites asking for clarification on the Yelp forums.**

In this study, we study the requirements for becoming a Yelp Elite. We explore the Yelp Challenge Dataset and train machine learning models for classifying Yelp Elites from non-Elites. From these models, we examine features which are predictive of Elite status and draw conclusions on their meaning relative to the official Elite criteria defined by Yelp.

***Index Terms*—Yelp Elite, Yelp Dataset Challenge, Yelp**

I. Introduction

Yelp Elite Squad, introduced in 2005 [1], is a private membership in Yelp. Elite members have ”Elite” badges bestowed on their Yelp profile and are invited to private parties held in local businesses [2]. Every year, local Yelp Elite council selects the upcoming year's Elite squadron among nominated (including self-nominated) and invited members in the council's local area. The membership expires after one year and the user has to be reselected to become the following year's Elite. Despite high number of enthusiastic Yelpers who want to become an Elite, Yelp officially does not disclose many details about their selection criteria – only high level description of three criterion – *Authenticity*, *Contribution*, and *Connection*. *Authenticity* is whether users use real name, a real profile photo, and an honest, unbiased opinion. Since the given dataset does not provide attributes related to this, we do not consider this for our prediction model. *Contribution* is about how much users have created meaningful contents in Yelp such as photos, reviews, tips, and more. The composite profile is reviewed meticulously by Elite council members. *Connection* is the number of vote (useful, funny, or cool) and compliments on user's activity. Since these two criteria are accessible to be quantified as a measure of Elite quality, we use attributes related to these criteria in our research as basic features. Yelp Elite Council members look for ”*je ne sais quoi* when reviewing Elite candidates... we know it when we see it” [3]. In addition to *Contribution* and *Connection*, we aim to identify 'hidden factors' which would be especially useful to strong candidates who are only missing *je ne sais quoi*.

II. Related Works

Although there is no previous work on discovering Yelp Elites, there are similar research done in the expert finding domain. One such research, by Zhang, Ackerman et al., explored methods for discover Java experts in online forums. By using the graph structure of user posts, they were able to develop modifications to graphical ranking algorithms based on PageRank and HITs call ExpertiseRank. This ranking algorithm took advantage of user behaviors on the forum, measuring the number of replies users have made or number of replies made to the user. They simulated their algorithm on a simulated Java forum network and showed that their ExpertiseRank showed a high correlation for identifying expert users [6].

Another expert finding study, by Zhang, Tang et al.[7], researched expert topic finding using graphical approach along node local information. For their research topic, they choose to study expert finding from a corpus of documents, given a target topic to rank experts by. They developed an algorithm which contains two steps, a local based initialization step and a second graphical expert score propagation step. In the first step, they used local author information to calculate an initial best guess expert score. This step computed the probability authors were an expert using simple statistical measures such as number of documents written in the target topic, or number of appearance as co-author. In the second step, a social graph is constructed based on relationships between authors. The social graph contains directional relationships such as co-authorship (bidirectional), personal relationships (directional), and advisor (directional) as edges. Each edge is weighted on the important of the relationship. Each author node is then initialized an expert score from the first step. Scores are then propagated for many iterations, summing the scores between nodes sharing edges. Zhang showed that the combined steps using local information initialization outperformed the graphical score propagation method alone.

III. Exploratory Analysis

A. Yelp Challenge Dataset

We obtain the dataset from the Yelp Dataset Challenge [4] which includes data from 10 cities. We use data from section 'business', 'review', 'user', and 'tip'. Each has the following fields:

1) business

business id, business name, neighborhoods, full address, latitude, longitude, average review star rating, total review count, categories, open status, hours of operation, attributes.

2) review

business id, user id, average review star rating, text, date, votes

3) user

user id, name, review count, average review star rating, total vote count, list of friends, years held Elite membership, yelp join date, compliments received, number of fans

4) tip

text, business id, user id, date, likes received

TABLE I
GENERAL STATISTICS OF THE DATASET.

Number of users	366715
Number of Elite users	25301
Number of reviews	1569264
Number of tips	495107
Number of business categories	783
Yelp join date range	2004-10, 2015-01
Review date range	2004-10-12, 2015-01-08

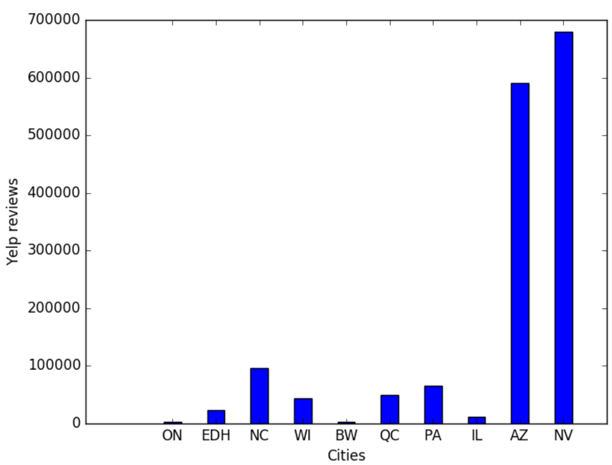

Fig. 1. Review count per region. We see that the majority of reviews are from Arizona and Nevada.

B. Elite Statistics

1) Selectivity

From the data, we see that the selectivity of Elite selection has increased over the years. The number of Yelp user and the number of Elite selection has both increased every year as shown in [Fig. 2] and [Fig. 3]. However, the ratio of Elite users to all users has a decreasing trend [Fig. 4].

2) Consecutive selection

From 2005 through 2015, we look at the occurrences of consecutive Elite selection of a user year to year [Fig. 5]. Not surprisingly, instance of no consecutive (n=1) selection is the most frequent, followed by a consecutive selection (n=2), three years in a row selection (n=3) etc.

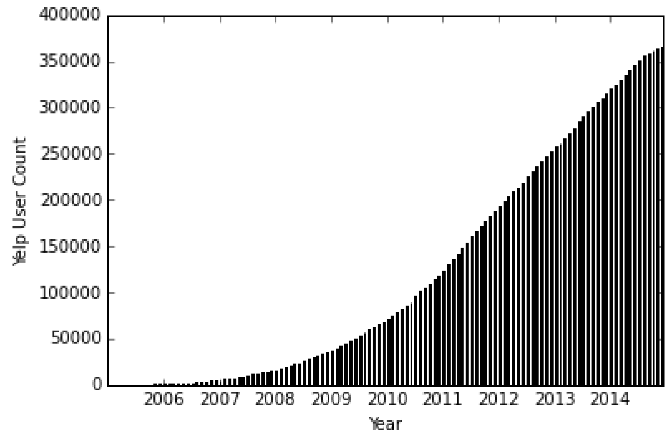

Fig. 2. Accrued User count joined in 2005-2015

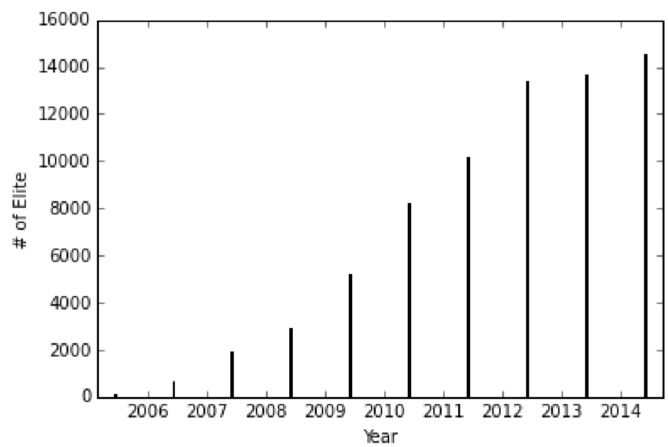

Fig. 3. Elite user count in year 2005-2015

IV. CORRELATION ANALYSIS

We use Pearson Correlation to explore relation between feature and Elite status. Using intuition, we choose the following features to analyze.

A. Content generation count

Metric	Pearson Correlation
Number of Review Count	0.2388
Number of Tip Count	0.0933

Number of review counts are more highly correlated than tip counts. Tip system was introduced later, which could explain the lower correlation. Difference between correlation of word count and length are small

B. Generated Content Statistic

Metric	Pearson Correlation
Average Review Word Count	0.1444
Average Review Length	0.1486
Average Tip Word Count	0.0126
Average Tip Length	0.0132

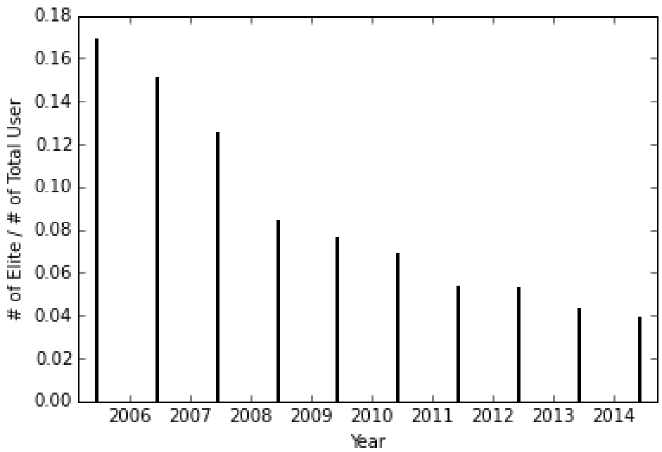

Fig. 4. Ratio of Elite user vs total user count each year in 2005-2015

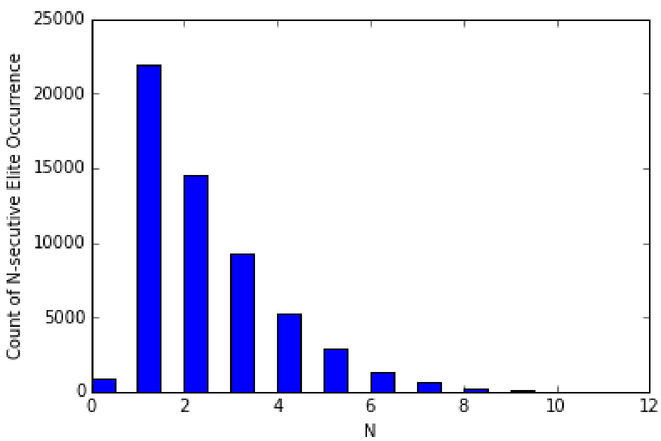

Fig. 5. Occurrences of N consecutive Elite selection among Yelp Elite 2005-2015

As in the previous section, quality of review text is more highly correlated than the quality of the tip text. .

C. User's Star Rating Trend

Metric	Pearson Correlation
Average Deviation of Rating	-0.0378
Average ABS Deviation of Rating	-0.0977
Average Review Rating	0.0162
Ratio of One Star Reviews	-0.0971
Ratio of Two Star Reviews	0.0050
Ratio of Three Star Reviews	0.1290
Ratio of Four Star Reviews	0.1323
Ratio of Five Star Reviews	-0.0953
One Star Review Count	0.0540
Two Star Review Count	0.1890
Three Star Review Count	0.2243
Four Star Review Count	0.2444
Five Star Review Count	0.1967

Count and ratio of four star ratings have the highest correlation.

D. Quality of Content judged by other users

Metric	Pearson Correlation
Total Review Funny Vote Count	0.1477
Total Review Cool Vote Count	0.1505
Total Review Useful Vote Count	0.1668
Average Review Funny Vote Count	0.2421
Average Review Cool Vote Count	0.2508
Average Review Useful Vote Count	0.1770
Total Review Vote Count	0.1586
Average Review Vote Count	0.2379
Average Tip Like Count	0.0780
Total Tip Like Count	0.0933

Average vote counts have slightly higher correlation than the total count. Average review of cool and funny votes have the highest correlation.

E. History

Metric	Pearson Correlation
Number of Days from Yelping Since	0.2840

F. Community Engagement

Metric	Pearson Correlation
Number of Friends	0.3332
Number of Fans	0.3478
Number of Elite Friends	0.3448
'profile' compliments Count	0.0579
'cute' compliments Count	0.0723
'funny' compliments Count	0.1812
'plain' compliments Count	0.1833
'writer' compliments Count	0.1991
'list' compliments Count	0.0466
'note' compliments Count	0.2120
'photos' compliments Count	0.0372
'hot' compliments Count	0.1919
'cool' compliments Count	0.2100
'more' compliments Count	0.1110
Total compliments Count	0.1728

In table [II, III], the correlation measure of number of friends, fans, and Elite friends show that the use of social features in Yelp increases chances of being an Elite. Receiving 'note', 'cool', or 'hot' compliments on the profile seems more correlated than receiving 'profile' or 'photos' compliments.

G. Review Business Category Diversity

In table, we show top six business categories in terms of relative frequency among reviews by Elite users and by non-Elite users. While more than half of non-Elite user's reviews are in 'Restaurant' business category, only twenty percent of the Elite user's reviews are. From this, we can hypothesize that Elite selection either favors 1) users who review more frequently in areas not reviewed by others, or 2) users who have more diversified review history in terms of business category. There are not enough data points to evaluate 1).

TABLE II

Business Category	Ratio in Elite Reviews
Restaurants	0.191
Nightlife	0.049
Food	0.046
Bars	0.042
Arts & Entertainment	0.029
American (New)	0.025
American (Traditional)	0.024

TABLE III

Business Category	Ratio in non-Elite Reviews
Restaurants	0.512
Nightlife	0.099
Food	0.090
Bars	0.088
American (New)	0.065
American (Traditional)	0.061
Mexican	0.052

We explore several metrics to featurize 2) and run Pearson Correlation against Elite status.

Metric	Pearson Correlation
Number of Categories	0.295
Range	0.156
Highest Ratio	-0.117
True Diversity Index	0.248
Simpson Index	-0.167

1) **Number of Categories**: Number of distinct business categories user reviewed.
2) **Range**: Ratio of Category with the Highest relative frequency - Ratio of Category with the Lowest relative frequency.
3) **Highest Ratio**: Ratio of Category with the Highest relative frequency.
4) **True Diversity Index**: $(\sum_{i=1}^{R} {p_i}^q)^{\frac{1}{1-q}}$ where R = number of categories [5]. We use q = 2 which yielded the highest correlation.
5) **Simpson Index**: $\sum_{i=1}^{R} {p_i}^2$

H. Locality

Yelp Elite applications are processed locally – after a user submits the application, the local community council reviews the candidate's profile and makes a collective decision. We suspect how much user contributes to Yelp on local businesses could be a strong feature in the Yelp Elite election process. In correlation to Elite status, we measure this 'locality' as the number of local Elite friends each candidate has. From [Fig. 6], we see that our dataset source from ten distinct cities world-wide and each city has enough number of data points for extracting locality feature.

We use K-means clustering algorithm (k=10) for user clusters. Since the dataset does not have any user location information, we estimate each user's current location by reviews written. We assume that location of businesses user write is highly correlated with user's location.

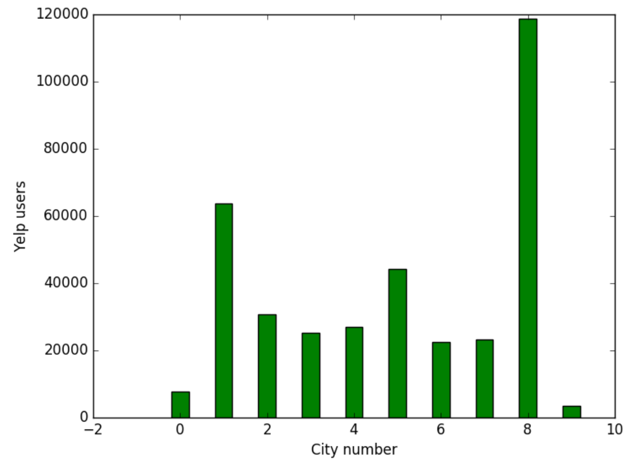

Fig. 6. The number of Yelp users in cities extracted by the K-means algorithm.

TABLE IV
STATISTICS ABOUT LOCALITY FEATURE

Name	Value
Average Local Elite Friend Count	1.5661
Average Local Elite Friend Count of Elite User	**29.9899**

Our result shows that Elite users have much higher average of number of local Elite friends than non-Elite users [IV]. This clues that locality factor has strong correlation with Elite status.

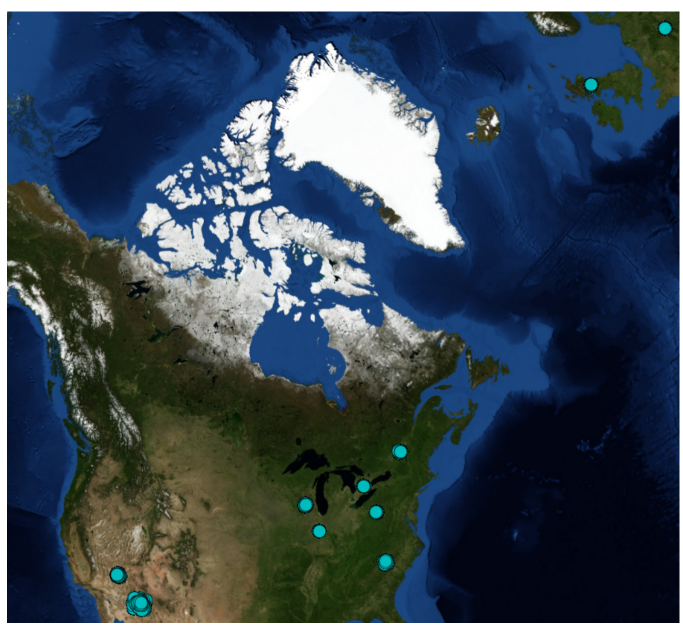

Fig. 7. Geospatial distribution of Yelp users

Metric	Pearson Correlation
Locality	0.294

V. Predictive Task

For our task, we predict if Yelp users are or have been Elite during their Yelp career.

A. Model Selection

We use logistic regressions and SVM as our predictors. In addition, we construct a friendship graph using friend list data from user profiles. We use PageRank and HITs to measure an authority score for each user. The authority score is used as a feature for our classifiers.

B. Feature Selection

We use features which show a high correlation result from our exploratory analysis. We add features to our model one at a time, prioritizing high correlation features first. Features are added until there was no difference in classification performance. The following features are added.

- Joined Year
- Number of Fans
- Fans to Friends Ratio
- Number of Friends
- Number of Elite Friends
- Elite Friends Ratio
- PageRank
- HITs Authority Score
- Compliment Votes to Fans Ratio
- Number of Reviews
- Number of Tips
- Number of Votes to Reviews Ratio
- Max Number of Reviews in one Year
- Number of Reviews by Year
- Number of Reviewed Business Categories
- Boolean for Reviewing in Category
- Number of Reviews by Stars
- Average Review Length
- True Diversity
- Boolean for Reviewing in City
- Number of Local Elite Friends

C. Results

We split data into three parts, 70% on train, 15% on validation, and 15% on test. The classification performance is shown in the following tables [V, VI].

We extract weights from our classifiers. On examination on the weights and their associated features, our logistic regression classifier appears to be generally fitted and our

TABLE V
Classification Results: Train

Classifier	Precision	Recall	F1 Score
Logistic Regression	0.860	0.715	0.781
SVM	0.875	0.704	0.781

TABLE VI
Classification Results: Test

Classifier	Precision	Recall	F1 Score
Logistic Regression	0.850	0.700	0.768
SVM	0.851	0.677	0.754

SVM classifier to be overfitted to the Yelp dataset. The logistic regressor rates general features such as ”elite to friends ratio”, ”compliments ratios”, and ”number of fans” very high while our SVM rated very esoteric category of places highly, such as ”castles”, ”radio stations”, and ”television stations”. We examine these esoteric features and find that they associate to places where very few people review, but very high ratio of reviewing users are Elite. For our feature analysis, we discuss features from the logistic regressor since they generalize better.

D. Features Analysis

In the trained model, correlating features can be examined in the following categories: Yelp socialness, compliments, category of places, cities, and reviews.

For Yelp socialness, the top influencing features is number of Elite friends ratio, followed by number of fans, and lastly number of local Elite friends. These features imply that number of Elite friends matters the most, and number of fans is more impactful for becoming Elite than number of friends. Unexpectedly, the PageRank and HITs authority feature do not impact the classification. We believe social features may already be captured by local profile information, especially by the ”number of Elite friends” and profile compliments features.

The top compliments feature are ”writer”, ”more”, and ”hot”. There is a large weight gap between these features and the next three positive compliments, which are ”cool”, ”list”, and ”note”. The most negatively weighted compliments are ”cute”, ”funny”, and ”photo”. It appears Yelp rewards Elite to users who contribute useful content to the website, with the ”writer” compliment being the most important factor. The high weight of the ”hot” compliment presents an interesting suggestion of some bias for attractiveness.

Fig. 8. Categories of Places Reviewed by Elites

Under categories of places reviewed, top features include ”Airports”, ”Burgers”, ”Dance Clubs”, and ”Breweries” [Fig. 8]. These features suggest that Yelp Elites are more outgoing and like to visit places of social gathering.

Fig. 9. Categories of Places Reviewed by non-Elites

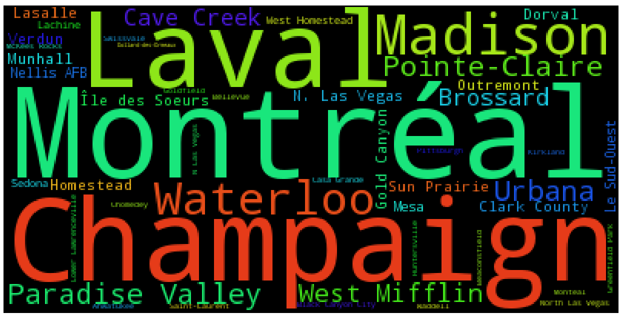

Fig. 10. Cities of Places Reviewed by Elites

Top features in cities [Fig. 10] are a mix of American and Canadian cities. Top Elite cities are ”Montreal”, ”Champaign”, ”Madison”, and ”Laval”. Negatively correlated cities [Fig. 11] are large, possibly diluting the overall population of Elites. They include ”Henderson”, ”Las Vegas”, ”Phoenix”, ”Edinburgh”, and ”Charlotte”.

For review features, we capture basic qualities such as review star ratings, votes, and number of reviews. These features have small positive weights relative to other features. For binned reviews by stars, the features rank highest from ”number of 4 star reviews” to the bottom at ”number of 1 star reviews”. This may imply that positive people are more likely to be Elites. Writing many negative reviews may be frowned upon when applying for Elite. In terms of votes, reviews with ”cool” and ”useful” are weighted higher than ”funny” reviews. In terms of number of reviews by year, those who review a lot in 2015 and 2004 are more likely to be Elite than have lots of reviews any other year, with number of reviews in 2007, 2008 and 2005 being at the bottom.

VI. Challenges and Future Work

During our implementation, we encounter a few challenges we need to address. Solving these challenges might improve our results. One aspect we do not concentrate on is extracting features from review text. We attempt to augment our features with unigrams from the review text, selecting terms with high tf-idf score. Unfortunately, using unigram terms do not increase our classification by much. We examine the unigrams weights and see that most of them were stop words, showing little relation to review sentiment or topic. One possible solution is to perform topic modeling on review text. Extracting topics from review text may capture aspects of Elite review text better.

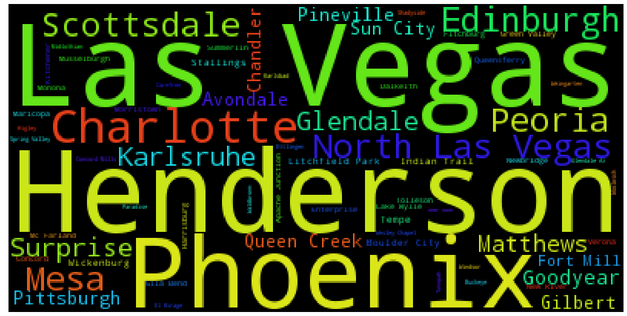

Fig. 11. Cities of Places Reviewed by non-Elites

Another challenge involves the lack of user profile data. The Yelp challenge dataset is mainly focused on reviews of businesses and omits use profile data. For example, there is no photo information or timestamped information. Specifically, there is lack of temporal data – when exactly users became Elite or information when users are elected as Elites–, and there is no information regarding social interactions between users; We do not know who voted or complimented whom. If we had these information, we expect our model to improve performance.

VII. Conclusion

For aspiring Yelp Elites, the selection criteria are often vague and confusing. Applicants are told to focus on criteria mentioned on the elite webpage, such as *Authenticity*, *Contribution*, and *Connection*.

In our study, we concretely discover which features correlate to Yelp Elite. From our results, users who write positive, useful and cool reviews are more likely to be Elite. Writing these reviews in turn leads to users receiving more compliments and fans, further increasing chances of becoming Elite. Those who write only funny reviews or negative reviews are less likely to pass the Elite application step. Yelp Elites tend to have more fans and Elite friends. Yelp Elite may have been more likely to receive nominations from these users, improving their chances during their Elite application.

In addition, we discover many qualities associated with Yelp Elite, though it may not be part of criteria of becoming Elite. We show that there is a strong correlation between Elites and outgoing personalities. We find Elites are more likely to visit socially oriented establishments, possibly enjoying an evening with friends at their local pubs or burger joint [Fig. 8]

We believe the attributes we discover aligns with Yelp's interest of accruing useful review (”Contribution”) and increase user engagement(”Connection”). These qualities are the hidden details of criteria Yelp uses to approve Elite applications.

TABLE VII
F1 SCORES OF MODELS ON YEAR 2007-2015

Year	07	08	09	10	11	12	13	14	15
07	*0.16*	0.31	0.27	0.26	0.22	0.2	0.17	0.13	0.08
08	0.33	*0.33*	0.32	0.32	0.3	0.28	0.26	0.25	0.25
09	0.12	0.27	*0.29*	0.29	0.26	0.23	0.21	0.18	0.15
10	0.18	0.32	0.3	*0.3*	0.28	0.26	0.25	0.22	0.2
11	0.14	0.31	0.31	0.3	*0.29*	0.27	0.27	0.25	0.18
12	0.12	0.28	0.31	0.31	0.29	*0.28*	0.28	0.27	0.31
13	0.16	0.32	0.31	0.31	0.3	0.29	*0.3*	0.29	0.28
14	0.09	0.23	0.29	0.3	0.28	0.28	0.29	*0.28*	0.3
15	0.04	0.12	0.16	0.17	0.2	0.2	0.25	0.27	*0.28*

APPENDIX A
YEAR-BY-YEAR MODEL

In this model, we aim to predict a particular year's Elite, given previous data. We do not use features that do not have temporal annotation (friends, fans, compliments, etc). With much less data, there is a limitation in performance of the model. However, the temporal model gives us clues to see how Elite criteria has changed over time.

A. Model

$$P(Elite, y) = \sigma(\theta \cdot [features_{y-1}, features_{year_{min} \cdots y-2}])$$

Yelp Elite selection process which occurs at the end of every year. Thus, the model creates separate features dedicated to year immediately before the target year.
We use logistic regression using the following features:
number of review count, number of tip count, average review length, average tip length, ratio of star ratings, average review votes, number of days yelping since, true diversity index

B. Result

This model performed better than $P(Elite, y) = \sigma(\theta \cdot [features_{year_{min} \cdots y-1}])$ (without features dedicated to previous year only).
The f1 scores of the model are in table [VII]. Each row represents target train year (trained with previous data), each column represents year tested against, and each cell contains the f1 score.

The f1 score correlates with abs(train year - test year), signifying that the criterion of Yelp Elite selection has changed over the years.

[Fig. 12], [Fig. 13], and [Fig. 14] show weight of features of the trained model that has increasing or decreasing trends over the years 2007 through 2015. Note that the input data was normalized from zero to one while training.

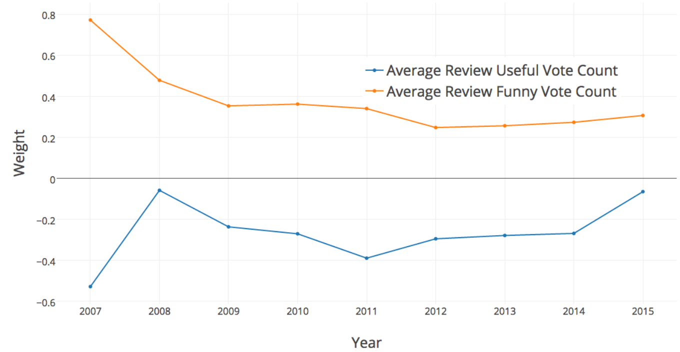

Fig. 12. Features with Decreasing Magnitude over the years 2007-2015

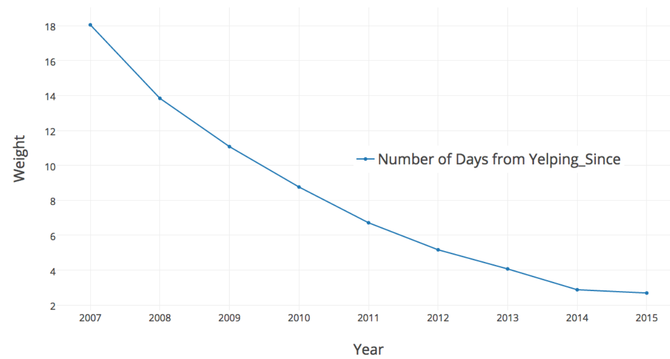

Fig. 13. Features with Decreasing Magnitude over the years 2007-2015

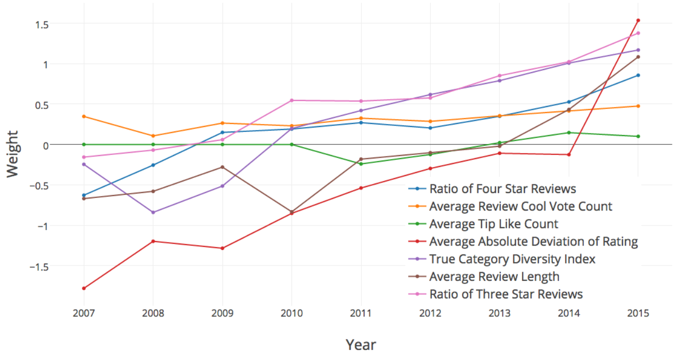

Fig. 14. Features with Increasing Trend over the years 2007-2015

REFERENCES

[1] J. Stoppelman (2009, March 5), Yelp Elite Events: What's the deal? [Blog]. Available: http://officialblog.yelp.com/2009/03/yelp-elite-events-whats-the-deal.html

[2] M. Stone (2014, August 22), Elite Yelpers Hold Immense Power, And They Get Treated Like Kings By Bars And Restaurants Trying To Curry Favor. Available: http://www.businessinsider.com/how-to-become-yelp-elite-2014-8

[3] Yelp, Yelp Elite Squad. Available: http://www.yelp.com/elite

[4] Yelp, Yelp Dataset Challenge. Available: http://www.yelp.com/dataset_challenge/

[5] H. Tuomisto (2010), "A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity," Ecography, 33: 222. doi:10.1111/j.1600-0587.2009.05880.x

[6] Zhang, Jun and Ackerman, Mark S. and Adamic, Lada (2007), "Expertise Networks in Online Communities: Structure and Algorithms," Proceedings of the 16th International Conference on World Wide Web, WWW '07: 221230. doi:10.1145/1242572.1242603.

[7] Zhang, Jing, Jie Tang, and Juanzi Li (2007), "Expert finding in a social network," Advances in Databases: Concepts, Systems and Applications. 1066-1069.