

Popularity Growth Analysis and Prediction on Yelp

Yilong Li, Zehui Wang, Zhouchangwan Yu
 Stanford University, Stanford CA 94305
 {yilong, wzehui, zyu21}@stanford.edu

Abstract—Research on the spread of information has been gaining popularity in recent years but most of the research has focused on social media platforms like Twitter and Facebook. In this project, we analyze the cascading behavior on Yelp, with a focus on how business gets popular. We analyze the relationship between business and users and how information distribution within user social networks influences the popularity of the business. Based on the analysis we then use GraphSAGE [6] framework to train embeddings for each business based on its cascading graph to predict its popularity.

I. INTRODUCTION

Social media has become an important source of information about a wide variety of businesses, and people have become more dependent on the information from social media when making consumer decisions. Social review websites such as Yelp serve as platforms for people to exchange their opinions about the businesses through reviews, ratings, photos, etc. Understanding the popularity growth based on the network between businesses and consumers is of great importance for business owners and platform service providers to make business and marketing strategies. In this project, we analyze the how user behaviors influence the popularity of businesses and make predictions based on the features extracted from our analysis. Specifically, we would like to analyze how business gets popular on Yelp and make predictions from its users.

We divide the work into two parts: **graph analysis** and **popularity prediction**. The goal of analysis is to find how a business gets popular - whether it grows gradually or explodes at a certain time period and how user behavior will influence the popularity of the business. Then we do prediction based on features that are extracted from the conclusion of our analysis¹.

II. RELATED WORK

A. Measuring User Influence in Twitter: The Million Follower Fallacy [1]

In this article Cha et al. discuss about the characteristics of "influential" users in social networks like Twitter. In Twitter the authors use three metrics to measure the influence of a certain user: indegree (number of followers), retweets (retweets of his posts) and mentions (replies to his tweets). And the authors reach the following conclusions:

(1) By analyzing the top influentials across these three measures, and measuring the correlation between different metrics for different groups, the authors find that the correlation between indegree and other metrics are quite weak, i.e. the most-followed person are not necessarily the most influential.

(2) From the spatial (topics) perspective, by analyzing the distribution and correlation for the metrics on tweets about different news topics, the authors find that for top influentials, there are more correlations over different topics, i.e. top influentials usually can have significant influence over a variety of topics, and the influence of users on different topics all follows the power-law trend.

(3) From the temporal perspective, the authors analyze the dynamics of influence over time. Different groups — top news agencies and celebrities — have different temporal characteristics for both retweets and mentions: For retweets, all groups have small increases of retweets over time, while for mentions which requires more interaction between twitter users and the "influentials", we can see a large decrease for top accounts (news accounts) and a large increase for the top 100-200 users which requires more self-advertisement and thus has more involvement with users.

B. Cascading Behavior in Yelp Reviews [2]

Khan et al. analyzed cascading behavior within Yelp user social network. Their work can be divided to two parts: structural analysis of cascades and cascade growth prediction.

In structural analysis, they summarize some frequent cascade topologies. The most common cascade topology corresponds to the case where a person participates in a cascade under the influence of only one other node. While this particular type of cascade represents more than fifty percent of the cascades across all the cities, statistics show that receiving the influence from more than one friends increases the likelihood of participation of the cascades.

Before the experiment of predicting cascade growth, they categorized the cascades as long or short cascades: If the length of a cascade is greater than or equal to the 90th percentile of the length of all the cascades of the city, then this is a long cascade; similarly a short cascade has a length that is less than the 90th percentile. Then the problem of cascade growth prediction is a supervised classification problem: predicting whether one cascade is long or short. They gathered features including root features (features of the original node who started the cascade), non-root features, business features and adopted gradient boosting as the learner. The results showed that the first reviewer may not be that importance in the case of cascades in Yelp reviews or in other words, non influential nodes may start the cascades.

C. The Tube over Time: Characterizing Popularity Growth of YouTube Videos [3]

Figueiredo et al. characterize the growth patterns video popularity on YouTube, the most popular video sharing ap-

¹The link to our code: https://github.com/zehuiw/yelp_popularity_analysis

plication. The two goals of their analysis are to understand (1) how the popularity of individual videos evolves over time since the video is uploaded (2) how users reached a given video by different types of referrers.

The authors collected data from YouTube statistics and divide into three datasets: (1) popular videos on top lists maintained by YouTube (Top); (2) videos that were removed due to copyright violation (YouTomb); (3) random sampled videos (Random)

To understand the popularity growth patterns, the authors performed analysis on the each dataset based on two aspects: (1) the time interval until the video reached most of its popularity (2) the burst of popularity experience by a video in days or weeks. They also further analyze the temporal dynamics of videos experience bursts activity by categorizing into Viral videos, Quality videos, and Junk videos. For the referral mechanism, the authors identified 14 types of referrals. For each dataset, they study the fraction of views from each category and the wait time until the first access from each category.

The authors conclude that the popularity growth pattern depends on the video dataset. The Top videos experience sudden bursts, while the copyright-protected videos experience a viral epidemic-like propagation. For all three datasets, search and YouTube internal mechanics are two most important referral mechanisms.

D. Critique

1) *Measuring User Influence in Twitter: The Million Follower Fallacy [1]*: In this article, by analyzing three common metrics of Twitter users, the authors successfully differentiate the users' popularity, which is measured by number of followers, and the users' influence over the network, which is more measured by retweets and mentions. The network is easy to acquire and the methods are all very straightforward — they focus on some top users and make statistics about their behaviors. There are also some deficiencies for this article: For the topic related characteristics, it lacks some categorized information, i.e. is there any difference in topics the top influencers focus on by user categories, since news media will have a much larger topic correlation factors than celebrities and other user-generated contents creators. Also the number of topics they research on is too small — they only have three topics while the timeline of a celebrity can be more complicated. For future work, since the temporal increase of users' mention influence for the emerging influential users is very clear, and we can use similar metrics within a fixed time window to predict new influential users in the future.

2) *Cascading Behavior in Yelp Reviews [2]*: This paper focuses on the cascading behavior among user social network in Yelp. The author analyzed the structural characteristics of cascades and predict how cascades will grow. Previously most of the research on cascades behavior focused on social platforms like Twitter, while this paper extend the research to Yelp and its user networks. The novel idea is that the authors define cascading behavior based on temporal information and user networks. They assume that the information of a particular

business flows through users' social network, i.e., only two users are friends will the information be cascaded. However, the mechanism of Yelp allows people to watch reviews of a business that are written by anyone, which means that information can be distributed without social networks.

3) *The Tube over Time: Characterizing Popularity Growth of YouTube Videos [3]*: This paper provides several metrics to characterize the popularity growth patterns of YouTube videos and the referral mechanism of them. One of the weakness of this paper is the method that the authors use to characterize the temporal dynamics of videos. They categorize videos into viral, quality, or junk videos, which is based on the fraction of views received on the most popular day. A more solid metrics is needed for the long-range popularity evolutionary patterns.

III. APPROACH

A. Dataset

Here we use Yelp Dataset Challenge [4] Round 12 in our project. It includes information about local businesses in 10 metropolitan areas across 2 countries.

For each business, we get all its basic information including business categories (restaurant, health care, cleaning, etc.), location (city, postal code, longitude and latitude) as well as its all review information, including the date, stars, text contents and the reviewer information for each entry, thus it will be possible to calculate current stars and ranking of a business at a certain time period.

B. Data Analysis

We analyze the dataset from three aspects: cascading behavior, business information and user-user network.

a) *Cascading behavior*: To understand how a business gets popular, we start with analyzing its cascading network. We first define that an edge (u, v) will be apart of a cascade for a business b if the user u writes a review or tip at time t and user v writes a review or tip at time t' such that u is the neighboring node of v and $t' > t$. Then for each business, we can generate its cascading graph and analyze its properties.

b) *Business information*: The goal of this part is understand the properties of the businesses on Yelp so that we can generalize a good metric to measure their popularity. We analyze business information from these aspects: locations, reviews (scores and rankings), spatial distribution, and temporal properties. Finally, based on our analysis we define a score to measure the popularity of each business.

c) *User-user network*: Since the cascades of a business depend on user's social network, we need to fully understand the user-user network on Yelp. We analyze its degree distributions, clustering coefficient and average ratings. We also analyze the differences between elite users and non-elite users in the network.

C. Prediction Model

After analyzing the properties of business, user network and cascading behaviors, we leverage the results we find to predict the popularity of a business. Based on our analysis

on user-user network, we generalize features for each user, i.e., their network degrees, whether they are elite users etc., as described in Section IV-C. In Section IV-A we also constructed cascading graph for each business. Then we adopt GraphSAGE [6], an inductive framework that leverage node features to generate node embeddings for unseen data. This framework learns a function that generates embeddings by aggregating information from a node’s local neighborhood. After generating embeddings of each node, we obtain the embeddings of the whole graph of the business by averaging all the individual node embeddings.

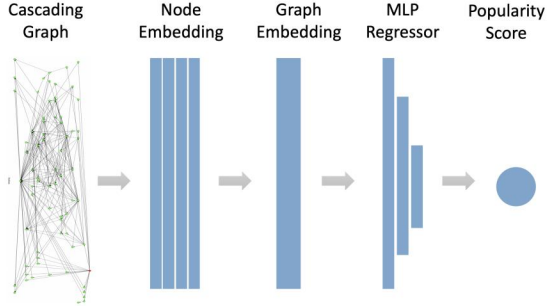


Fig. 1: Overview of the prediction model

The algorithm is described in Algorithm 1. We first initialize node embeddings to be input node features. Then at each iteration, nodes aggregate the embeddings of their neighbors and are combined with their previous embeddings. Finally, the combined embeddings are fed through a dense neural network layer and repeat the process. For simplicity, the aggregator function we use is the mean operation, where we take the elementwise mean of the vectors in $\{h_u^{k-1}, \forall u \in N(v)\}$.

After that we sum up the embeddings of each node to obtain graph embeddings, and then regress on the popularity as we defined in Section IV-B4 temporal analysis. We will use mean squared error as loss defined as

$$Loss = \sum (y_i - \hat{y}_i)^2$$

where y_i is the ground truth and \hat{y}_i is the prediction result. We use a multi-layer perceptron model as our regressor [7]. Fig. 1 displays the pipeline of our prediction model.

We will use R-squared (R^2) to evaluate our regression results. It provides a measure of how well observed outcomes are replicated by the model. Suppose our dataset has n values, then the ground truth popularity is denoted as $y = [y_1, \dots, y_n]$, with predicted values $\hat{y} = [\hat{y}_1, \dots, \hat{y}_n]$. Let \bar{y} be the mean of the data, then R^2 is

$$R^2 \equiv 1 - \frac{V_{res}}{V_{tot}} = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y_i - \bar{y})^2}$$

IV. DATA ANALYSIS RESULTS

A. Cascading Analysis

We first analyze the distributions of the length of cascades in different cities. We define the length of a cascade to be the longest path from the center node, which is the business here, to any leaf node. We find that most of the cascades are

Algorithm 1: Embedding generation algorithm. Adapted from GraphSAGE [6]

Data: Graph $G(V, E)$; node features $\{x_v, \forall v \in V\}$; depth K ; weight matrices $W^k, \forall k \in \{1, \dots, K\}$; non-linearity σ ; differential aggregator functions $Agg_k \forall k \in \{1, \dots, K\}$; neighborhood function $N : v \rightarrow 2^V$

Result: Vector representations z_v for all $v \in V$

$h_v^0 \leftarrow x_v, \forall v \in V$

for $k = 1, \dots, K$ **do**

for $v \in V$ **do**

$h_{N(v)}^k \leftarrow Agg_k(\{h_u^{k-1}, \forall u \in N(v)\})$

$h_v^k \leftarrow \sigma(W^k \cdot CONCAT(h_v^{k-1}, h_{N(v)}^k))$

$h_v^k \leftarrow h_v^k / \|h_v^k\|_2, \forall v \in V$

$z_v \leftarrow h_v^K, \forall v \in V$

small with maximum length for most cities below 10. The majority of business do not have cascades or have very short cascades but long cascades do happen. This can be explained by user-user network analysis later in Section IV-C, which shows that the node degree in user social network follows a heavily tailed distribution with the majority of users on Yelp have very small node degree. Fig.2 shows the distribution of the size of cascades in some large cities. The patterns follow a power-law distribution. Fig.3 shows some examples of review cascades of business across different cities. For each city we pick the business which has the longest cascading length. From Fig.3 we can see that some users have strong influence and spread the information to many other users.

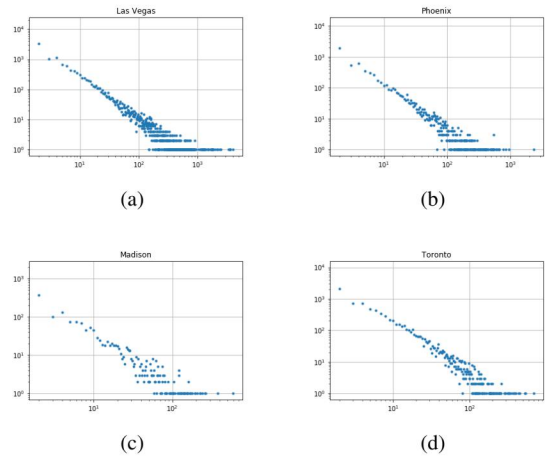


Fig. 2: Distributions of the size of cascades across different cities. X-axis shows the number of nodes that participate in the cascades. Y-axis is the number of business that has the corresponding size.

B. Businesses

1) *Cities:* In our dataset, there are totally 188593 businesses and 5996996 reviews. The data was collected in mainly 10

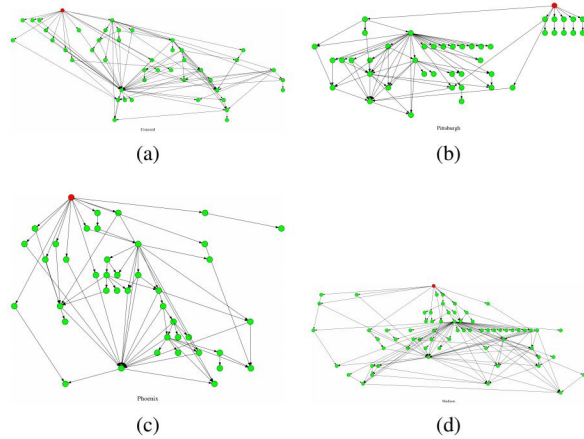


Fig. 3: Examples of review cascading of business. Green nodes represent users. Red nodes represent the business and its econet represents the users that start the cascading.

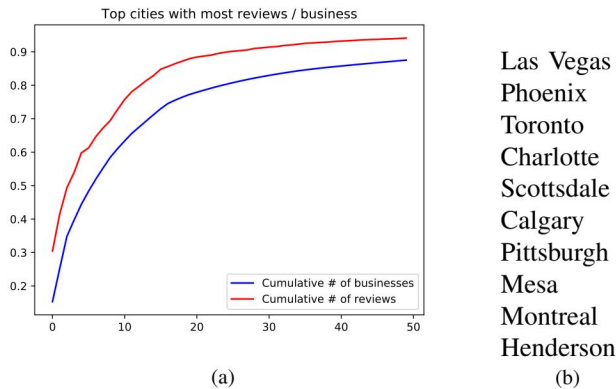


Fig. 4: (a) Cumulative number of businesses and reviews of the top cities (b) Ten top cities

metropolitan areas in US and Canada, and most of the data was around the center cities, including Las Vegas, Phoenix, Toronto, Charlotte, Calgary and Pittsburgh. About 76% of reviews and 62% of the businesses belong to these top cities, as shown in Figure 4, so in future analysis of business popularity, we will mainly focus on these cities.

2) *Reviews, scores and rankings*: For each review a user gives to a business, it contains a rating score ranged from 1 to 5 and an optional comment message, and the score of the business is just the average of all the rating scores it receives. The distribution of businesses and distribution of reviews grouped by their scores is shown in Figure 5. Here we can clearly see that, users tend to give 4 or 5 stars in most cases, and most of the businesses are rated as 3 to 4.5 stars. Note that most 5-star businesses have only a few comments, usually less than 20, and we filter these businesses out for our analysis. If we look into the statistics of reviews counts, we can find that the review counts for businesses generally follow the power-law distribution, as shown in Figure 6.

3) *Spatial distribution*: Since cities have different sizes and different layouts, the distribution of businesses can be

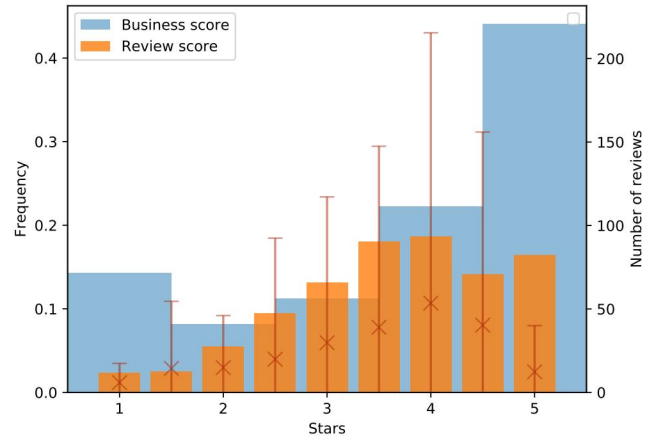


Fig. 5: Distribution of business score and user review scores

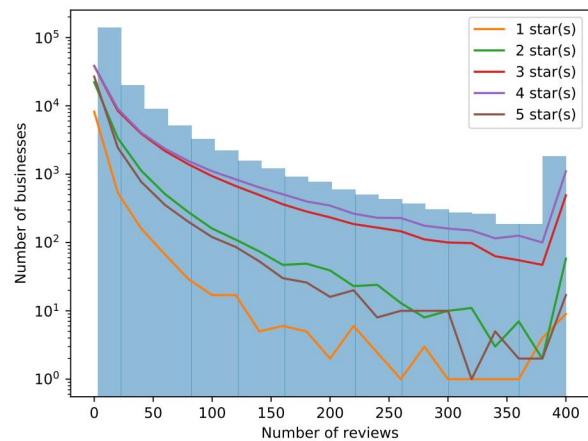


Fig. 6: Review count for businesses follows the power-law distribution

completely different. We can plot the distribution of businesses with different numbers of neighbors within a certain radius, as shown in Figure 7 — Here we can see that businesses in most cities are scattered or they form into small clusters with less than 150 nodes inside it (e.g. Las Vegas and Phoenix); there are also cities whose business district are more connected and more close to each other, like Montreal.

In order to look into the closeness of popular businesses, for the top-rated and most-reviewed businesses, we dynamically create a network of businesses based on the relative distance between nodes. We choose businesses with review counts or score greater than a given value as nodes of the network, and then add edges between nodes within a certain radius and evaluate the clustering coefficient as the radius increases. Here we take Las Vegas as an example. We can see that for all valid businesses the clustering coefficient reaches its maximum value at about 800 metres; for businesses with more than 100 and 200 reviews, the maximum value is reached at 1000 and 1200 meters respectively. Top businesses are mostly scattered

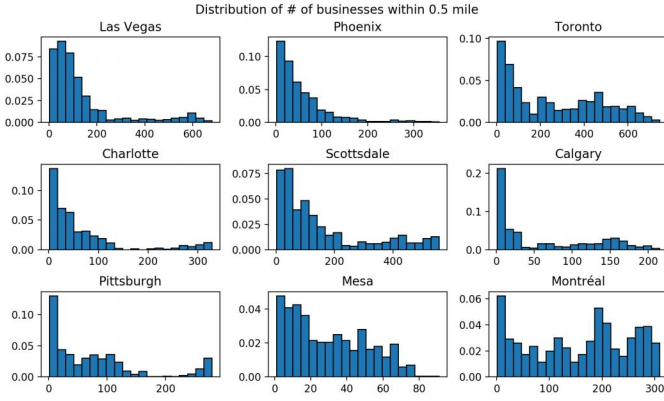


Fig. 7: Distribution of businesses with neighbors, horizontal axis: number of neighbors in 0.5 mi, vertical axis: percentage

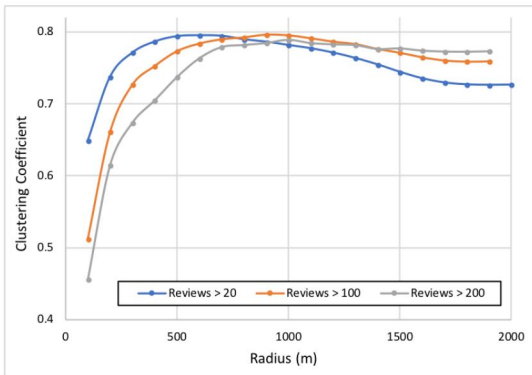


Fig. 8: Clustering coefficient of the spatial connection graph

rather than spatially connected closely.

4) *Temporal analysis*: Similar to the methods in [3], we can analyze the popularity (i.e. review count) growth patterns in our dataset by measuring the following metrics: (1) the time it takes for a business to get popular; (2) the time popularity (number of reviews) burst happens. Here we take Las Vegas as an example.

a) *Time taken to get popular*: For each business, we can calculate how many weeks it takes for the business to reach a certain percentage (e.g. 10%, 50%, and 75%) or a certain number (e.g. 100, 200) of reviews, as shown in Figure 9. We calculated the top 200, 1000 and 5000 rated businesses in Las Vegas, and we plot the time it takes to reach certain percentages of reviews. Unlike videos [3] or tweets [1] which get their maximums very quickly over the Internet, the number of reviews usually increase gradually and it becomes really hard (usually takes about 8 to 10 years) to get most of its current reviews. And we can clearly find that about 40% of the businesses reach the first 10% and first 25% of their reviews much faster than others, and this tendency holds for all different business subsets.

b) *Popularity bursts*: Here for each business we find the month with most reviews (“peak month”), of which the result is shown in Figure 10, and we find that for most businesses the peak month is on the first month after its opening — the first month is critical for businesses to get most of its initial reviews (usually 25 ± 20 reviews). And we also notice that for

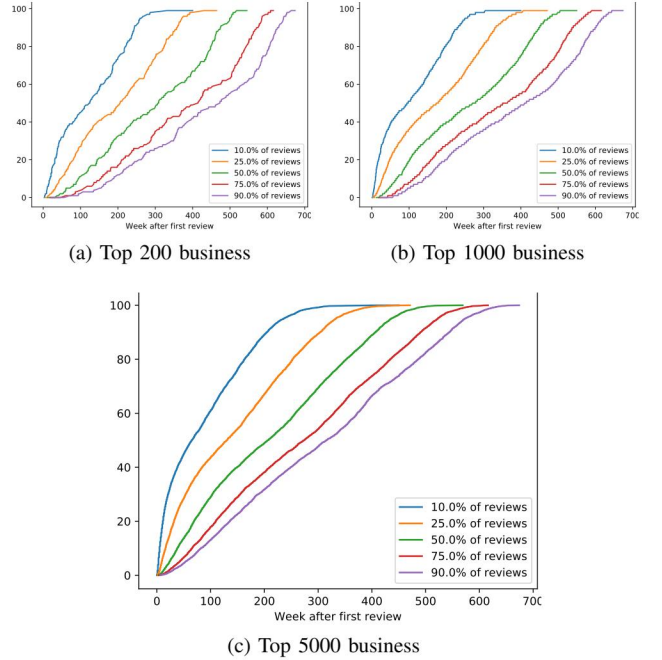


Fig. 9: Time taken for businesses to reach certain percentages of reviews. horizontal axis: number of weeks taken, vertical axis: percentage of businesses which reached that percentages of reviews

businesses with more reviews, this time can be later. Possible reason can be that it was recommended by some influential users or influential social medias, and these businesses will be what we focus on for future analysis stages.

C. User-User Network

There are 1,518,169 user nodes in the user-user network, in which 67,109 are elite users, and 879,891 users have at least one friend on Yelp. As shown in Fig.11, both all-users and elite-users generally follow a power-law degree distribution, $P(k) \sim k^{-\gamma}$. For elite users, γ is much smaller at low degrees and increases at high degrees. Fig.12 shows the cumulative clustering coefficient of all-users and elite-users. More than 70% users in the entire user group have 0 clustering coefficient, and the cumulative fraction of user increases rapidly as clustering coefficient increases. On the other hand, only 10% in the elite user group have 0 clustering coefficient, and the cumulative fraction of users increases slower. The statistics of average degree and average clustering coefficient is summarized in Table.I. Elite users tend to have more friends and cluster more closely.

For both all-user group and elite-user group, the average ratings of the users are shown in Fig.13. For the all-user group, the majority of average ratings have medium values, while there are two noticeable peaks at two ends, which means that some users only express one of the extreme feelings on Yelp review, i.e. 5 stars or 1 star. For the elite-user group, the average ratings follows a normal distribution with center at 4.0, which means that elite users tend to give more reasonable ratings with various degrees of preference.

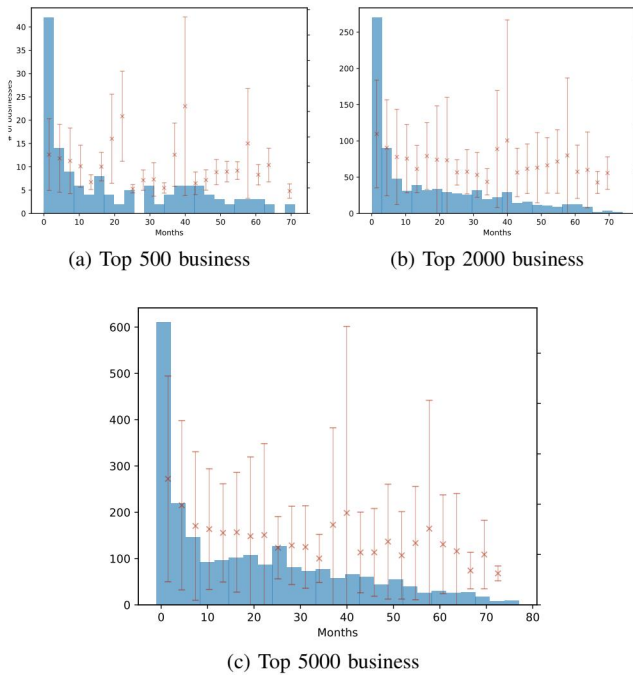


Fig. 10: Distribution of the “peak month” of businesses. horizontal axis: month, vertical axis: number of businesses which has peak month at this month

TABLE I: Statistics of all users and elite users

	Average Degree	Average Clustering Coefficient
All users	8.72	0.0431
Elite users	77.3	0.172

To understand the role of elite users, we randomly pick some elite nodes and draw the egonets of them, as shown in Fig.14. We find that for the elite users with very large egonets, the majority of the neighbors are elite users (red), and there are many edges between neighbor elite users. Even for elite users with smaller egonets, the neighbor elite users are more likely to be connected than non-elite users (blue). This is consistent with our analysis above that elite users have higher degree and clustering coefficient. Elite users are more socially active in the Yelp network.

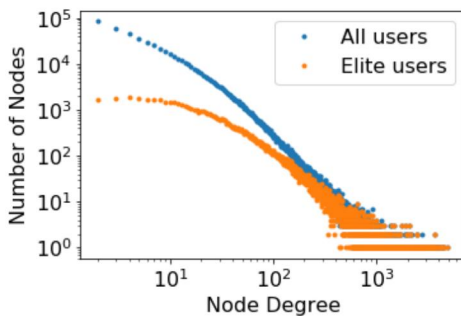


Fig. 11: Degree distributions of all users and elite users.

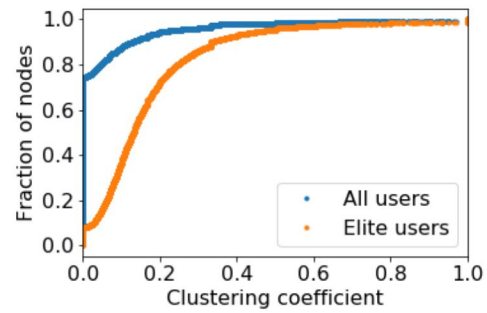


Fig. 12: Cumulative clustering coefficient of all users and elite users.

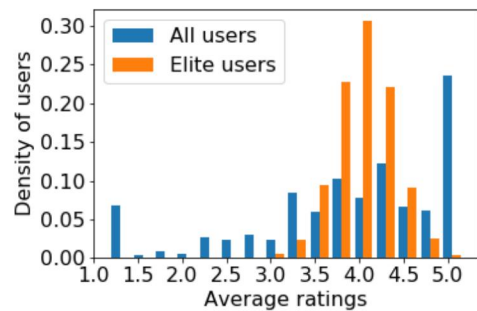


Fig. 13: Distributions of average rating of all users and elite users.

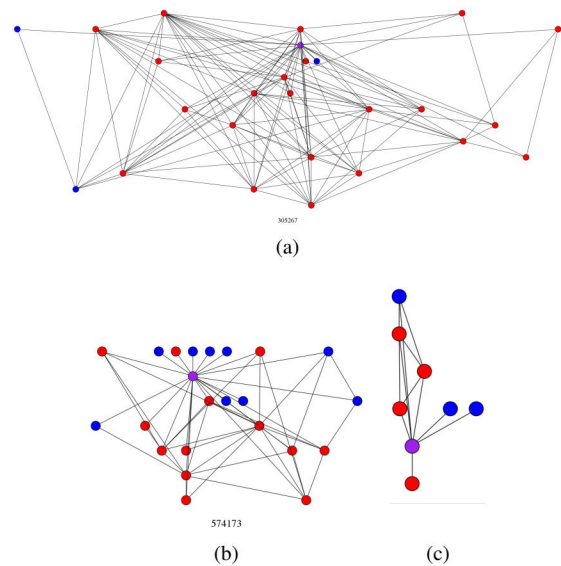


Fig. 14: Example egonets of elite users. Purple node represents the center elite user of the egonet. Red nodes represent neighbor elite users. Blue nodes represent neighbor non-elite users

V. POPULARITY PREDICTION EXPERIMENT

A. Implementation Details

We run experiments on the Yelp dataset to predict the popularity of a certain business.

For each business, we give one score to indicate its popularity. Since Yelp dataset only provides the number of reviews and average stars provided for a single business, we define the popularity of a business based on its relative ranking of reviews in the given area, i.e.

$$p(\text{business}) = \frac{\text{ranking of business within 2 miles}}{\# \text{ of businesses within 2 miles}}.$$

We first train the embedding of each business cascading graph using the approach described in Section III-C. The aggregator we use is mean operation. We initialize the feature vector of each node in the graph with our prior knowledge about the user: total counts of reviews, number of friends, number of useful/funny/cool votes sent by the user, the number of fans the user has, average stars, and the numbers of various complement types received. The resulting embedding is a vector of length 20 for each node. Then we take the mean of all node embeddings, and the sum of all node embeddings to get the graph embedding, which is the corresponding business embedding we use as the input of our regression model. We tried different regression models, including polynomial ridge regression (linear model), multi-level decision tree and SVM regression using radius-based Gaussian kernel. In our final model we use a multi-layer perceptron, i.e. fully connected neural network, as the regression model.

We compare the prediction result with that of a baseline model. The baseline model takes an input of manually-selected features from business data. For each business, its features include its location, the number of stars it gets, the number of reviews it receives and its opening time. Then we feed the features into the same regression model and do the prediction.

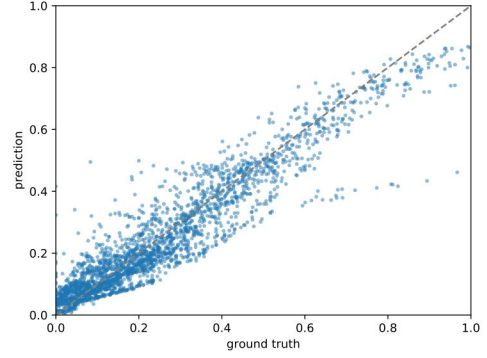
B. Result and error analysis

We use the R^2 correlation coefficient to analyze our result and the ground truth value, and compared different features we used in our regression models. We chose 10960 businesses randomly from the dataset (using only businesses in the top cities and with more than 20 reviews), dividing them into training, validation and testing set with proportion of 60%, 20% and 20%. The results is shown in Table II.

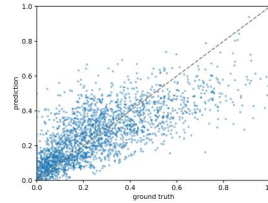
TABLE II: Comparison of different features used for popularity score regression

Feature	Baseline	Mean	Sum	Mean-Sum	Sum-Stat
R^2 -score	0.888	0.105	0.441	0.476	0.515

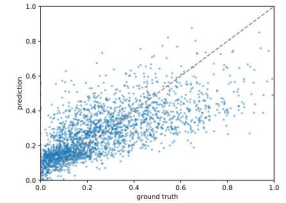
We can find some correlation between the predicted score using only cascade graph representation, though the correlation is still weaker than that with using hand-selected features. Using different features leads to different correlation values. We found that the features we learned from summation works the best with adding some graph statistics data into the features. There are a lot of reasons which caused the difference between our model and the baseline model:



(a) Baseline



(b) Sum-Stat



(c) Mean-Sum

Fig. 15: Scatter plots of output scores and ground-truth popularity scores using different features

1) *Node embedding*: Here we set the dimensionality of each node to be 20, which can be sometimes not enough for cascade graphs with many nodes (especially for those popular businesses) and complex network structures. The dimensions of the node embedding can be higher for future experiments.

2) *Graph embedding*: In our method we only calculated the sum and the average of all nodes inside the graph, which is a coarse estimation of the graph embedding. Other graph embedding methods should also be tried in order to get more accurate representation of the graph, e.g. calculating node embedding of an extra node connecting to the whole graph, or calculating the node embeddings of different random walk paths.

3) *Definition of popularity*: It is hard to define popularity on a large dataset like the Yelp one, since popularity itself is dynamic and has some locality (related to location and number of businesses nearby). Our currently definition of business popularity consider the relative ranking of businesses within 2 miles, while there are some problems when using this definition: (1) For high-ranking nodes we care more about its ranking (for example, the top 10 restaurant should have similar scores no matter compared within 100 businesses or 1000 businesses). (2) For low-ranking nodes we more care about its relative ranking since the absolute ranking value are not so useful. So we need a scoring function which can synthesize both relative ranking and absolute rankings.

4) *Cleanness of dataset*: Here we use a dataset including all types of businesses and we only filter the data by the location (whether it is located in the top ten cities) and number of reviews (we only choose businesses with more than 20 reviews). Since the cascading in different types of business

can be different, using only a certain type of business (for example, restaurants) will

VI. CONCLUSION

In this project, we aim to explore how a business gets popular and how it is reflected by user behaviors. We use Yelp Dataset Challenge Round 12 as our dataset and deliver the following things:

- 1) Graph analysis: we define the cascading behavior in the growth of a restaurant and analyze the properties of cascading graphs. In order to understand how a restaurant gains its popularity, we also deliver an temporal analysis of popular businesses on Yelp — how they get popular and how it is related with their interaction with influential and ordinary Yelp users. Last, we analyze the properties of user social network to investigate how they influence the popularity of the business.
- 2) Popularity prediction: We introduce a metric (popularity score) to measure the popularity of a business quantitatively. We then propose a popularity prediction algorithm which is based on cascading graphs. We compare the R^2 results of our proposed model and baseline model and provide detailed analysis.

REFERENCES

- [1] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and P. Krishna Gummadi. 2010. Measuring User Influence in Twitter: The Million Follower Fallacy. Proceedings of ICWSM (International Conference on Web and Social Media) 2010.
- [2] Muhammad Raza Khan. 2017. arXiv:1712.00903. <https://arxiv.org/abs/1712.00903>
- [3] Flavio Figueiredo Fabricio Benevenuto Jussara M. Almeida. 2011. The Tube over Time: Characterizing Popularity Growth of YouTube Videos. Proceedings of the fourth ACM international conference on Web search and data mining. Pages 745-754. <https://dl.acm.org/citation.cfm?id=1935925>
- [4] Yelp, Inc. Yelp Dataset. <https://www.yelp.com/dataset/challenge>
- [5] Jure Leskovec, Ajit Singh, and Jon Kleinberg. 2006. Patterns of influence in a recommendation network. In Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, 380389.
- [6] Hamilton, Will, Zhitao Ying, and Jure Leskovec. "Inductive representation learning on large graphs." Advances in Neural Information Processing Systems. 2017.
- [7] Glorot, Xavier, and Yoshua Bengio. "Understanding the difficulty of training deep feedforward neural networks." Proceedings of the thirteenth international conference on artificial intelligence and statistics. 2010.

INDIVIDUAL CONTRIBUTIONS

Zehui: Defined and analyzed cascading behaviors; Trained embeddings of cascading graphs; Implemented the baseline model.

Zhouchangwan: Did analysis on user-user network; Prepared node features for prediction model.

Yilong: Did analysis of businesses and users based on Yelp dataset; Implemented the business scoring and prediction part from node embeddings of cascade graphs; Maintained the server and database.