

Application of Network Methods for Yelp Review Recommendations

Adharsh Murali, Andrew Barber, Koki Sasagawa, Phil Robinson

University of Michigan School of Information

SI608

### **Research Question and Motivation**

When using Yelp, users generally look up specific businesses by title or search by category. Upon making a search, the application will return a list of relevant local businesses that are rated on a five-star scale. Once a business is selected, additional information is provided which may include images of dishes and reviews left by other users. As a crowd-sourced platform, these reviews are an essential source of information for users to consider when making a decision about a business. Currently, yelp implements an algorithm called the "Yelp sort" which utilizes a combination of users elite statuses, review recency, and review popularity to sort the reviews.<sup>1</sup> The customer can also manually sort based on recency, popularity, or elite status. This approach, however, fails to take into consideration the similarities and differences that exist between individual users. For example, a review left by a user who averages 4-star ratings may not be as relatable to a critical reviewer who averages 3-star ratings. Our team is interested in exploring whether reviews can be sorted based on the similarities of user's rating trends and business preferences against other users. Several network-based models are used to see if a network structure can capture similarity and differences among users, and effectively sort reviews in an order that is meaningful to a user.

### **Related Work**

In general, there are three types of recommendation systems: 1) Collaborative filtering 2) Content-based filtering and 3) Knowledge-based filtering.<sup>2</sup> Collaborative filtering methods leverage user-item interactions (in our case, ratings that users give to businesses).<sup>2</sup> It is based on the assumption that users are similar to each other if they like similar items and rate these items

similarly. Content-based filtering uses information describing items (i.e., business type) to define object distance between items.<sup>3</sup> Items similar to users *past* preferences would be presented as recommendations.<sup>2</sup> Lastly, knowledge-based filtering systems utilize *active* user specifications to compute recommendations.<sup>2</sup>

Given these options, we ultimately decided on using collaborative-based filtering to define user similarities. Collaborative filtering methods have been used for a variety of applications including recommendation systems for Flickr photo groups,<sup>4</sup> YouTube video suggestions,<sup>5</sup> and question recommendations produced by the world's largest Chinese question and answer site.<sup>6</sup> The project by Jeff Han, Justin Kuang, and Derek Lim shared many procedural similarities to our own project as they used collaborative filtering as a means of defining user similarity. Given a user and a restaurant, they found a group of similar users using cosine similarity, then, using a combination of ratings similar users have assigned to the restaurant to predict what a user will rate the given restaurant.<sup>7</sup> Another study by Kritika Singh attempts to predict user rating based on collaborative filtering methods.<sup>8</sup> The evaluation method used in Singh's study is similar to our evaluation protocol and thus should be noted. As neither of these studies utilizes network methods in their approaches, we aim to explore the applicability of network-based methods in order to predict user reviews and optimize review-recommendation.

### **Data**

The dataset for this project was adapted from the Yelp Dataset Challenge:

<https://www.yelp.com/dataset/challenge>. We utilized three separate files for our analysis:

business.json, user.json, and reviews.json. The data structures for these files can be seen in

*Figure 1* below. The original size of this dataset included 188,593 businesses, 440,865 users and 5,996,996 reviews.

Implementing network-based methods using a dataset this size proved to be problematic due to high computational costs. As such, we decided to reduce our data down to query-able instances based on location and business type. Businesses that were closed and did not have valid state codes were removed from the dataset. Additionally, businesses were limited to those that were strictly categorized as restaurants. We believe filtering the data this way successfully allows us to group data that share meaningful characteristics (e.g., keeping categories commonly shared among restaurants while removing those that are not). Additionally, our filtering reflects how Yelp filters search results to be localized to the location of the user and return results that match their query.

We finalized our sample by only including restaurants from Arizona labeled as “Chinese”. After filtering the size of the data includes 33,575 users, 538 businesses and 47,359 reviews. This represents a unique use-case that we believe is representative of a real-world utilization for our review-recommendation system.

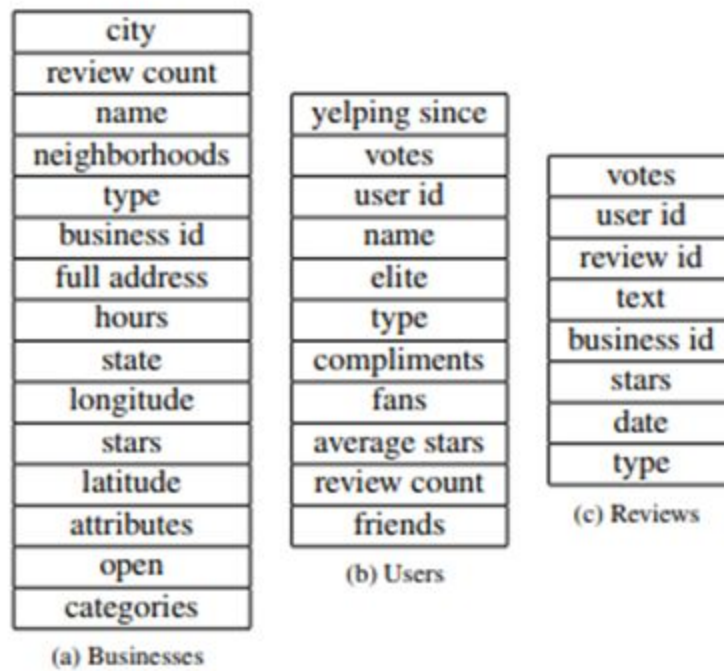


Figure 1: Structure of Yelp dataset. Reprinted from  
 "Predicting user rating for Yelp businesses leveraging user  
 similarity". Kritika Singh.. 2015

## Methods

### Overview

As mentioned above, we decided upon a personalized collaborative-filtering method as a means to define user similarity. We reasoned that users who reviewed the same business would be more similar to each other - this was the basis behind our definition of similarity. In terms of networks-based concepts we had two components to our rationale: 1) In a user-user graph where edge weights are determined by the number of businesses the two users both reviewed, the two nodes are similar if they have a low shortest path length and there are many paths that connect

them. 2) In a bipartite graph connecting users to categories of business with edge weights based on the user's average rating of that business category, users who share common neighbors with each other will be more similar. The first component is the rationale behind our implementation of the distance-based method and the eigencentrality method. The second component is the rationale behind our implementation of the Jaccard similarity method and the Adamic/Adar method.

### **Networks Used**

After we finalized data filtering, we created networks to be used in our methodologies. We created three graphs: 1) User-to-business bipartite graph with weighted edges representing individual reviews of that business 2) User-category bipartite graph connecting users to American, Italian, Indian and Mexican restaurant categories with edges representing the average rating a given user gives that category (the average category score for each user in the user-category bipartite was taken from the dataset pre-filtering by state and category) 3) User-user one-mode projection of the user-business bipartite graph with edge weights representing the number of businesses connected users have both reviewed. The structure of these graphs are visualized in *Figures 2-4* below.

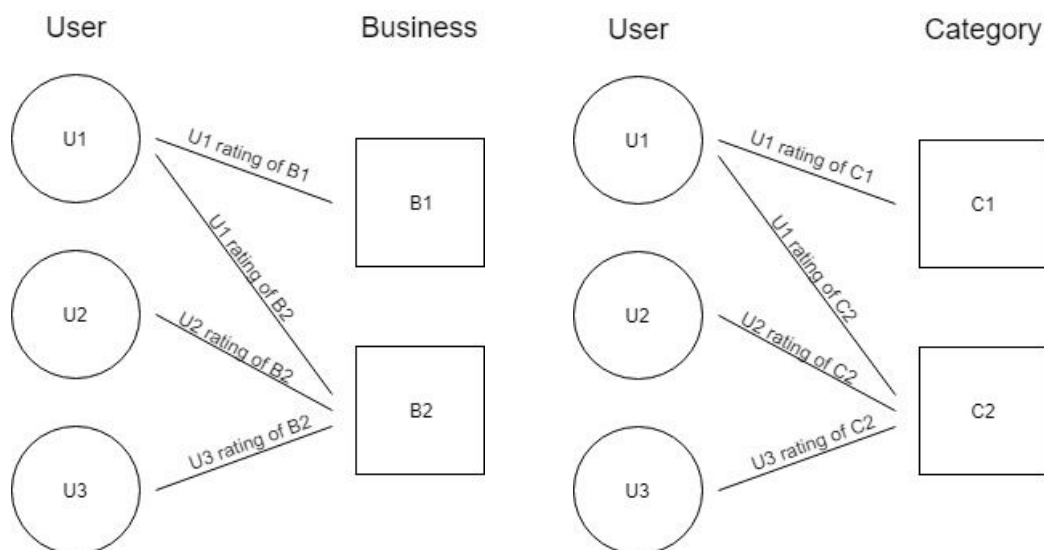


Figure 2: Structure of user-business bipartite graph

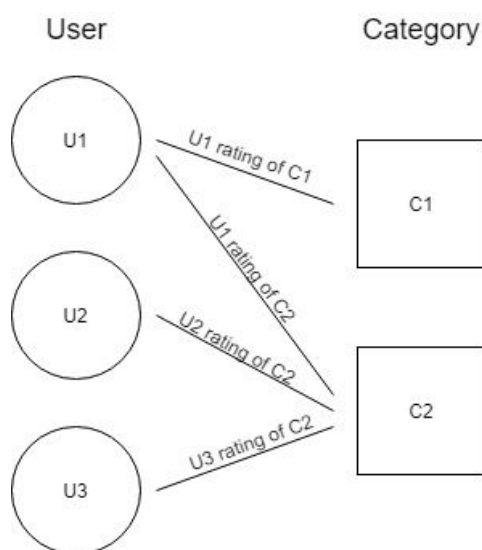


Figure 3: Structure of user-category bipartite graph

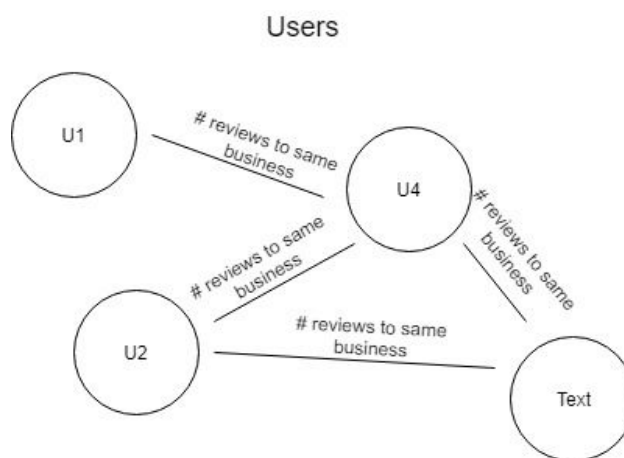


Figure 4: Structure of one-mode projection of user-business bipartite graph

### Distance-Based Approach:

The intuition behind this approach is that two nodes are similar if they can reach each other with a minimum number of steps and also have multiple simple paths to reach each other. In the project, using the user-business bipartite network, we built an ego network of all the users

who have given reviews to the queried business. Using the Breadth-first traversal, the shortest path between the querying node and the users in the ego networks are computed. The users are ranked based on nodes which are very close and also similar to the querying node. To evaluate we used the average ratings of top 10 reviewers similar to the querying user and assigned the user a most probable rating that the user might give for the restaurant.

### **Eigenvector Centrality**

The Eigenvector centrality-based rating prediction works on the hypothesis that the rating of the node depends on the most influential node in the network. In the user one-mode projection of user-business bipartite network, the EigenVector centrality was computed based on the number of connections a node has in the network. Taking advantage of the fact the network was highly connected, for a new querying user, the user with high Eigenvector centrality in the querying user's neighborhood who has given the rating to the restaurant was inherited to give the rating for the restaurant the querying user is interested. The user reviews were ranked based on the EigenVector Centrality. To evaluate the performance of the ratings of the querying user to the restaurant was computed based on the average rating of the top 10 user rating similar to the querying user.

### **Jaccard Similarity & Adamic/Adar**

With the creation of the user-business and the user-category bipartite graphs, we can query a test tuple, containing a user id and a business id. With the business id, we take the ego



network of the user-business bipartite to find all reviewers who have left a review of this restaurant. Having this, we know which users we will want to measure against in the user-category bipartite graph. The idea for using the user-category graph is that users who have more rated categories in common will return a higher similarity score. We take the nodes that all have the highest similarity score and then find a good way to inherit the ratings of these similar nodes. Because the graph is about 83% sparse (as in 17% of all total edges in the user-category bipartite graph exist), a higher similarity score indicates that two users have many common categories in common. Since the selected categories are of mostly unlike nature, we try and create more sparsity in the user-category graph. For example: if two categories are “American” and “Steakhouse”, it wouldn’t be too far of a stretch to assume that a user rating a steak house would also rate an American restaurant, as these categories have a significant overlap in food style. If the two categories are much more dissimilar, then common sense would dictate that these two categories won’t necessarily draw similar clientele, and thus, give much different similarity scores for neighbors when we query the user id, business id query tuple.

## **Baseline**

We used two baselines for our results. The first baseline is based on the average of all the ratings the target user has given all businesses throughout their time on yelp. This baseline suggests how well the ratings users gave in the past are at predicting how they will rate in the future. The second baseline is based on the average of all ratings given to the target business. This suggests how well the average rating of all other users predicts how any given user will rate a business.

## Results

In order to evaluate our methods, we used each method to compute the ratings of user-to-business reviews. We then took the root mean square error of these predictions compared to the true ratings of the reviews. This method is very similar to that used by Kritika Singh in her study “Predicting user rating of Yelp businesses leveraging user similarity”.<sup>8</sup> As there is no established ground truth for user similarity, we used rating-prediction as a proxy based on the assumption that similar users would rate the same business similarly. The results of our RMSE can be seen in *Figure 5* below.

| Method                                  | RMSE |
|---|------|
| Baseline 1 (Average rating of business) | 1.26 |
| Baseline 2 (Average rating of user)     | 1.19 |
| Jaccard                                 | 1.77 |
| Adamic/Adar                             | 1.60 |
| Distance-based                          | 1.27 |
| EigenCentrality                         | 1.35 |

*Figure 5: RMSE results for our similarity-ranking methods*

### **Challenges**

We faced two major problems with the dataset, which we did our best to overcome. The first challenge was that this data was not very well constructed to be a network, thus making a network-based solution ambiguous. We have enough data to create an edge-edge list from the data, but there are many different ways to define what the network could consist of. This isn't a problem in of itself inherently, but does become a problem within challenge two: the sheer size of the data set. When we look at creating a user-user edge list of users who have rated the same restaurant, a complete graph of a restaurant of 6000 reviews (we saw many examples of this magnitude), creates more than 17 million edges. This dataset simply was not meant to be used in its entirety as a standard one mode projection network.

### **Future Work**

One of the main issues of Collaborative Filtering methods is the sparsity of data or the cold start problem where users have a lack of information that makes it difficult for the algorithm to reliably predict a rating. The entire yelp dataset of users and business reviews was very sparse. It may have been interesting to consider other sources of information, such as business metadata to create a business x category graph where the relationship is dependent on shared features and not explicit links.

### **Conclusion**

Based on our results, methods that relied on defining user similarity based on the specific businesses users each reviewed outperformed methods that relied on defining user similarity based on general categories. Methods that utilized the user-user one-mode projection network

(distance-based and eigencentality) fared much better than those that utilized the user-category bipartite graph (Jaccard and Adamic/Adar). From this result we can conclude that defining user similarities based on similar tastes in business category was not as effective as defining user similarities based on similar businesses reviewed.

What is more, the baseline results which predicted ratings based both on the average rating of the target business and the average rating of the target user each performed better than any of our methods. This means that for our subset of data, predicting ratings based both on the user's rating history and the business's rating history is better than predicting based on user similarity. If true, then predicting a user's rating based on their past ratings and predicting what a business will be rated based on how other have rated it are both sound rating-prediction methods.

## References

- 1) How is the order of reviews determined? (n.d.). Retrieved from [https://www.yelp-support.com/article/How-is-the-order-of-reviews-determined?l=en\\_US](https://www.yelp-support.com/article/How-is-the-order-of-reviews-determined?l=en_US)
- 2) [Aggarwal, C., C. \(2016\) Recommender Systems: The Textbook, Doi: 10.1007/978-3-319-29659-3\\_1](#)
- 3) Dolgikh, D. and Jelinek, I. (2016) Graph-based Rating Prediction using Eigenvector Centrality. *Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 1, 228-233. Doi: 10.5220/0006044902280233
- 4) [Zheng, N., Li, Q., Liao, S., Zhang, L. \(2010\), Which photo groups should I choose? A comparative study of recommendation algorithms on Flickr, Journal of Information Science, Doi: 10.1177/0165551510386164](#)
- 5) [Baluja, S., Seth, R., Sivakumar, D., Jing, Y., Yagnik, J., Kumar, S., Ravichandran, D., Aly, M. \(2008\). Video suggestion and discovery for youtube: taking random walks throuh view graph. 17th International Conference on World Wide Web, 895-904](#)
- 6) Liu, Q., Chen, T., Cai, J., Yu, D. (2012) Enlister: baidu's recommender system for the biggest chinese Q&A website. *Sixth ACM conference on Recommender System*, 285-288. Doi: 10.1145/2365952.2366016
- 7) Han, J., Kuang, J., Lim, D. (2014) Predicting Yelp Ratings From Business and User Characteristics, *Stanford Universiy Course CS229 Machine Learning*, retrieved from <http://cs229.stanford.edu/proj2014/Jeff%20Han,%20Justin%20Kuang,%20Derek%20Li>

[m,%20Predicting%20Yelp%20Ratings%20From%20Business%20and%20User%20Characteristics.pdf](#)

- 8) Singh, K. (2015). Predicting user rating for Yelp businesses leveraging user similarity

[http://jmcauley.ucsd.edu/cse255/projects/wi15/Kritika\\_Singh.pdf](http://jmcauley.ucsd.edu/cse255/projects/wi15/Kritika_Singh.pdf)