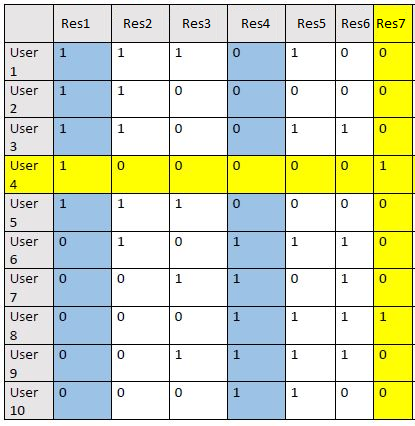
**Find Your Foodie-mate**

1. **Introduction**

Individuals in social networks are often unaware of people who share similar tastes, literally. From our own personal experience, we believe that people often bond over food, and so we want to recommend people of common food preferences and food-related activity. To do this, we aim to conduct an analysis of the Yelp network. To do so, we want to use Yelp user, business, and review data and find similar ratings between users. Our approach takes into consideration the Yelp user base, as well as user activity: creation of user reviews, ratings, cuisine and follows. Realistically speaking though, communities are much more likely to form when the members are in close physical proximity with each other, so our metric will also be sensitive to check-ins to the same restaurants, which gives us geographical data. The following proposal evaluates recommendation system. Advantages and disadvantages discussed range from the ability for an algorithm to capture belonging in multiple communities, to the computational viability of certain approaches for certain datasets.

* **Background**

We select some of popular restaurants of Chicago city from yelp whose followers belong to some ethnicity. We use this information to label the data and train a model. We call the popular restaurants as prime restaurants. First we came up with set of prime restaurants; we use these prime restaurants to label the data. Choosing these prime restaurants is very important, because these prime restaurants help us to learn other features of ethic group. The prime restaurants should be popular among particular ethnic group at-least 90% of followers who follows those restaurants should belong to same ethnicity and these prime restaurants should cover wide area set of audience. Of course it is difficult to come with the set of prime restaurants. With domain knowledge about that ethnic group it is possible. Once we came with prime restaurants, we sample 20 restaurants and 539 users of those restaurants. And then we construct a matrix (refer fig below) with users in y axis and restaurants they follow on x axis.



**II. Data Collection**

We made use of a web scraper called BeautifulSoup to scrape a user's profile page provided by Yelp to form our data set (refer fig below). The range of businesses is limited to locations in the area of Chicago, IL.



**III. Methods**

After we collect the data, we exported the data as a csv file. This makes our processing faster. Then we read through each record, we used python Counters and Dictionary to create list of dictionaries. Each element in the list represents row of our matrix and dictionary represent mapping from column to value. Then we used DictVectorizer to convert it in to sparse matrix.

**IV. Experiments**

Once we process the data we then fit norm and cosine similarity. Once this is done we next compare it with Jaccard and Pearson correlation.

**V. Related work**

[Yelp recommendation system](http://cs229.stanford.edu/proj2013/SawantPai-YelpFoodRecommendationSystem.pdf). In this paper the approach used is Clustering and k-fold cross validation. Our approach is unique as we have used recommendation system.

**VI. Conclusions and Future Work**

We are using a static data set, so we cannot model how the network changes over time. With real-time data, we could use composite scores to inform optimization--since internal representation is a large complete graph, it would be computationally expensive to update the composite score frequently. Rather, we could revisit each edge (and recalculate the composite score between two users) at a frequency proportional to the composite score.

An interesting extension could be to label the clusters defined by the model. The clustering algorithms use edges to form clusters but do not give us reasons behind this clustering, hence we can provide specific names or ‘labels’ to each food taste cluster. We can determine the labeling of clusters through methods such as looking at the single most popular restaurant category loved by the users in the community.