The Effect of Network Measures and Geolocation on Business Ratings

Eirini Kapogli
Stevens Institute of Technology
School of Business
Hoboken, NJ
ekapogli@stevens.edu

Harris Kyriakou
Stevens Institute of Technology
School of Business
Hoboken, NJ
ckyriako@stevens.edu

SUMMARY

User reviews are increasingly affecting how people make choices (Luca and Zervas 2013) and how organizations operate. Online platforms can be used to examine how network as well as physical location attributes may affect users' ratings over time. We analyzed reviews submitted in Yelp, the largest business review site to date. We examined the longitudinal effect of network centrality measures to business ratings reported by users. In addition, we examined the effect of business geolocation attributes not only to the network status of the business, but also to its reviews. Our preliminary results show that specific clusters of physically collocated businesses attract high reviews, whereas other clusters receive more negative reviews. Our findings indicate that reviewers are more likely to vote for one of the two extreme review values over time. We suggest that such a behavior is a result of the reviewers' effort to streamline the reported quality of a business with their own perception. We also find that people are more likely to provide a review for businesses ranked in the third quartile of all business. We discuss implications of our findings and future directions for practitioners and research.

INTRODUCTION

Yelp is an online "urban guide" and business review site. The website uses a 1-5 star-based rating scale. Users can provide reviews or tips, check-in on businesses or upload photos through mobile applications. Yelp has incorporated many social networking features such as friendships and likes. Many businesses incentivize their customers to provide reviews on Yelp by providing discounts and free products. This collective reviewing system creates a uniquely dynamic business knowledge corpus, where the wisdom of the crowd decides which business will prevail.

Issues such as review fraud (Luca and Zervas 2013; Mukherjee et al. 2013), predicting product ratings by utilizing information present in reviews (McAuley and Leskovec 2013) and whether consumer reviews are good for business (Blanding 2011) have been addressed in the past. Scholars have suggested that on average the first ratings that businesses receive on Yelp overestimates their eventual reputation (Potamias 2012). Our analysis build upon this work and extends this findings, suggesting that reviewers are more likely to vote overrate or underrate businesses over time. The purpose of this

research is to identify how such a growing community evolves, to identify existing clusters, to identify the most important users and businesses based on various criteria and lastly to evaluate customers' review tendencies and qualities.

METHOD AND DATASET OVERVIEW

The dataset used was released by yelp.com and included 15,585 businesses from the greater Phoenix, AZ metropolitan area. Our dataset was longitudinal and included metrics from 70,817 users contributing 335,022 reviews. The data include the geographic coordinates, ratings and reviews of all businesses, social connections between users and dates that users visited each of the locations. The dataset was downloaded in a .json file and the RJSONIO (Lang 2012) package was used in order to import the data in R. the network was created with the igraph (Csardi and Nepusz 2006) and Gephi (Bastian et al. 2009) was also used for the visualization.

In order to efficiently examine the relationships between users and businesses, the need to transform the bipartite network into two different one-mode networks was identified. The first one contained only user nodes while the second one only business nodes. In the first business graph, two businesses would be connected to each other when a user had reviewed both of them. In other words, the business network is a co-citation (Small 1973) representation of the original bimodal network. Similarly, the user network was created through the bibliographic coupling method (Kessler 1963). An edge was created between two users when both of them had reviewed the same business. Both of the newly formed graphs contained edge weights representing the frequency that such a connection was observed.

NETWORK CENTRALITY ANALYSIS

Understanding the connections between users and businesses within a competitive environment will provide a two-fold contribution. Companies may use such insight to customer segments that are more likely to shop from them. At the same time, platforms connecting businesses and customers like Yelp can understand the inner mechanics of such processes to improve their suggestions, ultimately benefiting customers.

RATING ANALYSIS

In order to better understand how people interact with businesses, we analyzed the rating patterns of people over time. Our analysis shows an active and steadily growing community (Figure 1A). In addition, we find that businesses don't follow a normal distribution and most of them have a rating of 3.5-4 stars in a 5-point scale (Figure 1B). It is also important to notice that people are more likely to give a 1 out of 5, or a 5 out of 5 review over time (Figure 3C).
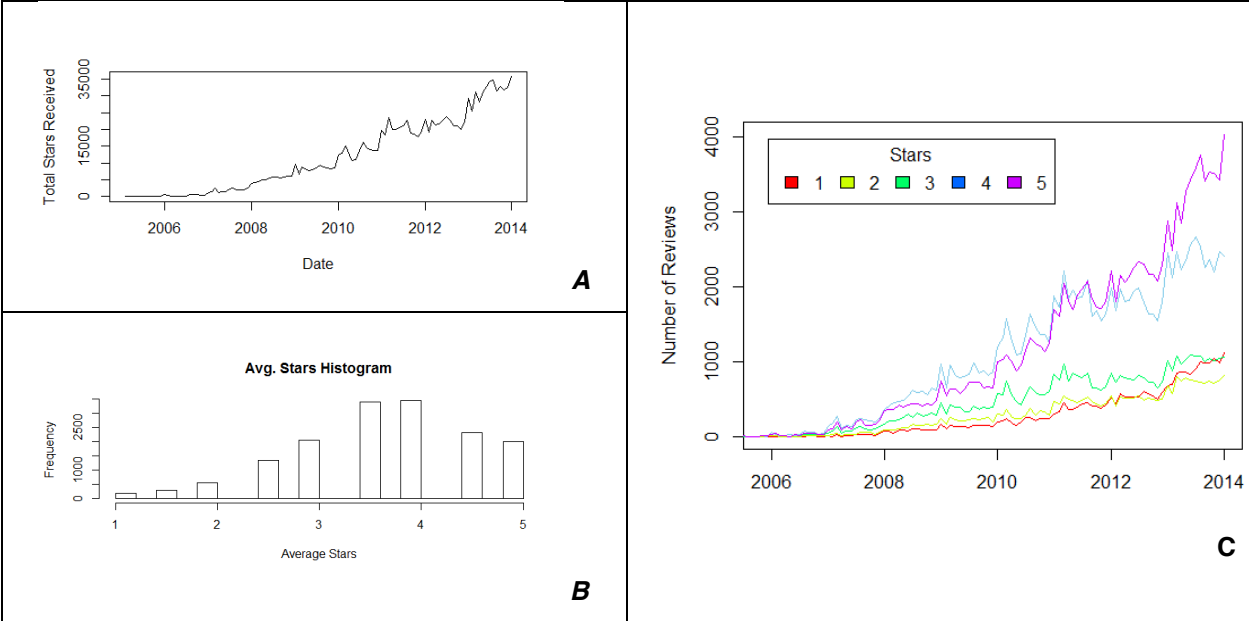
Figure 1. Figure 1A shows the total number of stars received by all businesses since 2006, suggesting a vibrant and growing community. Figure 1B shows that most businesses had an average score between 3.5 and 4 stars (not a normal distribution). Figure 1C shows that as Yelp attracts more reviews, users have the tendency to rate more extremely. The number of reviews is constantly increasing. Such a phenomenon can be explained by both the increasing number of businesses on Yelp as well as the increasing number of users. However, 1 star and 5 stars reviews are increasing with a significantly higher rate.
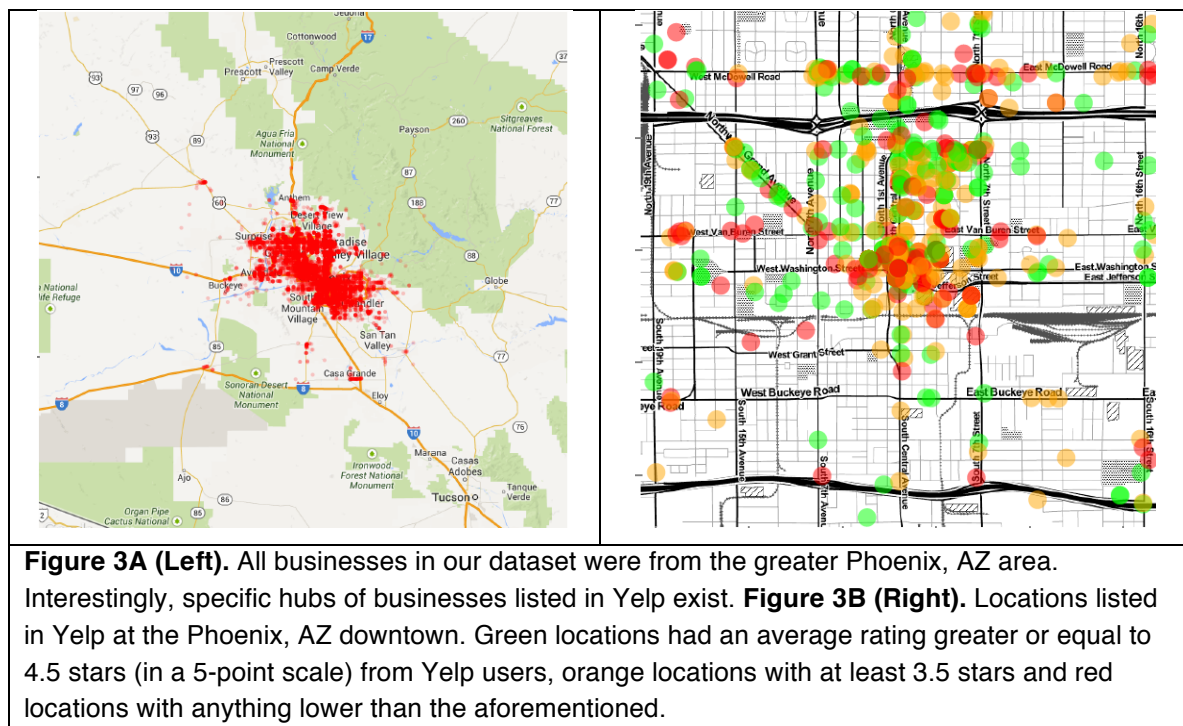
Next we examined which businesses were more likely to receive reviews by Yelp users. Our analysis suggest that businesses that good (but not very good) businesses, meaning that they were ranked in 3rd quartile of all businesses, were more likely to receive a review (Figure 2).

Figure 2. Average star rating and number of reviews received. Our analysis suggests that people are more likely to provide a review for businesses that are rated in the 3rd quartile of all restaurants. Error bars show standard error of the mean.

LOCATION ANALYSIS

Using geographic coordinates of all business, we created a plot indicating their position and their rating, hoping to understand the effect of competition to a company's success (Figure 3). Being close to a cluster of highly ranked businesses may benefit a company, as people are more likely to visit the area. In extreme situations, for instance when one of the incumbent competitors cannot accommodate all potential clients, dissemination of demand may happen. On the other hand, proximity to a very successful competitor or a set of competitors may create higher expectations to potential customers, resulting in a perception and review bias. Our future work will try to combine quantitative and qualitative methods to understand how these strategies affect the success of businesses.

Figure 3A (Left). All businesses in our dataset were from the greater Phoenix, AZ area. Interestingly, specific hubs of businesses listed in Yelp exist. **Figure 3B (Right).** Locations listed in Yelp at the Phoenix, AZ downtown. Green locations had an average rating greater or equal to 4.5 stars (in a 5-point scale) from Yelp users, orange locations with at least 3.5 stars and red locations with anything lower than the aforementioned.

REFERENCES

Bastian, M., Heymann, S., Jacomy, M., and others. 2009. "Gephi: an open source software for exploring and manipulating networks.," *ICWSM* (8), pp. 361–362.

Blanding, M. 2011. "The Yelp factor: Are consumer reviews good for business," *Harvard School of Business*.

Csardi, G., and Nepusz, T. 2006. "The igraph software package for complex network research," *InterJournal, Complex Systems* (1695:5), pp. 1–9.

Kessler, M. M. 1963. "Bibliographic coupling between scientific papers," *American documentation* (14:1), pp. 10–25.

Lang, D. T. 2012. *RJSONIO: Serialize R objects to JSON, JavaScript Object Notation*.

Luca, M., and Zervas, G. 2013. "Fake it till you make it: Reputation, competition, and Yelp review fraud,"
    Harvard Business School NOM Unit Working Paper (14-006).

McAuley, J., and Leskovec, J. 2013. "Hidden factors and hidden topics: understanding rating dimensions
    with review text," in *Proceedings of the 7th ACM conference on Recommender systems*, ACM,
    pp. 165–172.

Mukherjee, A., Venkataraman, V., Liu, B., and Glance, N. S. 2013. "What yelp fake review filter might be
    doing?," in *ICWSM*.

Potamias, M. 2012. "The warm-start bias of yelp ratings," *arXiv preprint arXiv:1202.5713*.

Small, H. 1973. "Co-citation in the scientific literature: A new measure of the relationship between two
    documents," *Journal of the American Society for information Science* (24:4), pp. 265–269.