Beer Hipsters: Exploring User Mindsets in Online Beer Reviews

Group 27: Yuze "Dan" Huang[*], Aaron Lewis[†], Garrett Schlesinger[‡]

December 10, 2012

1 Introduction

1.1 Motivation

Online product rating systems are frequently used to generate recommendation systems to boost sales. However, these systems do not take into account how user reviews are influenced by previous reviews of the same product. We wish to model how users interact with previously posted reviews. Namely, we want to see if there are distinct categories of users who tend to agree with previously posted reviews of an item, tend to disagree with previously posted reviews of an item, tend to only review items that have not been frequently reviewed, or exhibit one of these behaviors for a period of time, and as they spend more time in the community, switch to another behavior type. A thorough study of how users on a review site react to previous content will be useful as a psychological study with potential marketing applications. For example, if a user is known to agree with previously posted reviews, it is likely that they will positively react to any recommended product with many good reviews.

1.2 Dataset

Our work deals with reviews on the beer rating website `BeerAdvocate.com` [2]. This dataset contains over 1.5 million reviews from over 33,000 users on over 66,000 beers collected over the course of 10 years. Beers have as attributes their "style" (e.g. American IPA, English Stout) and the brewery that brews the beer. Each user review of a beer includes the following features:

- Numeric scores on five different aspects (appearance, aroma, palate, taste, overall)
- A timestamp indicating the review time
- A plaintext review provided by the user.

For the purposes of our project, we ignore the plaintext review, and in order to simplify our analysis and models, we only consider the "overall" aspect of each review score.

For convenience, we refer to the set of users as U, the set of beers as B, and the set of reviews as R. For a particular user u who has rated beer b, $R(u, b)$ is the review score written by u of b, $R(u, \cdot)$ is all review scores user u has written, and $R(\cdot, b)$ is the all review scores written for beer b.

[*] `yuze@stanford.edu`
[†] `amlewis@stanford.edu`
[‡] `gschles@stanford.edu`

1.3 Problem Definition

Our aim is to confirm the existence of "bandwagon" behavior in the `BeerAdvocate` dataset and present demographics of the network based on the extent to which each user "bandwagons". We consider "bandwagon" behavior to be the tendency of users to interact with beers based on either the positivity or volume of response by other users to the beer. One way to quantify this sort of behavior is a metric that we term the "bandwagon score." Specifically, a user's bandwagon score should be some combination of the correlation of their reviews to the mean review of a beer. We define a measure of similarity β between a single user u's rating of a beer b and the rating distribution of that b prior to be rated by that user,

$$\beta(u,b) = \frac{1}{2} - \mid \frac{1}{2} - cdf(R(u,b)|mean(R(\cdot,b)), std(R(\cdot,b)) \mid \tag{1}$$

where the mean and standard deviation in the formula are for ratings of the beer that occur before u rates the beer, and cdf is given $X \sim norm(\mu, \sigma^2)$. This measure assumes that reviews by user and by beer are both approximately normally distributed, a fact verified below. This metric is designed such that the closer u's rating is to the mean rating for b, the higher β will be. Figure 1 shows a plot of bandwagon similarity vs. score for a fixed mean and standard deviation. Figure 2 shows the distribution of aggregate bandwagon scores across all reviews of a user.

1.4 Hypotheses

We hypothesize that users who do not show high bandwagon score will tend to review beers that not many people have reviewed. The intuition here is that these users are exploring beers about which little is known, presumably in order to discover "hidden gems" in the world of beer. We refer to such users as "beer hipsters" and define the following "hipster coefficient" to quantify the extent to which a user u reviews beers that have only sparsely been reviewed before:

$$h_l(u) = \frac{\mid \{b : b \in B \wedge R(u,b) \in R(u,\cdot) \wedge |R(\cdot,b)| < l\} \mid}{\mid R(u,\cdot) \mid} \tag{2}$$

That is, the hipster coefficient of a user u is the ratio of the number of beers they review which have fewer than l reviews at the time of rating to the total number of beers they review. We expect that a high hipster coefficient will correlate to low bandwagon score.

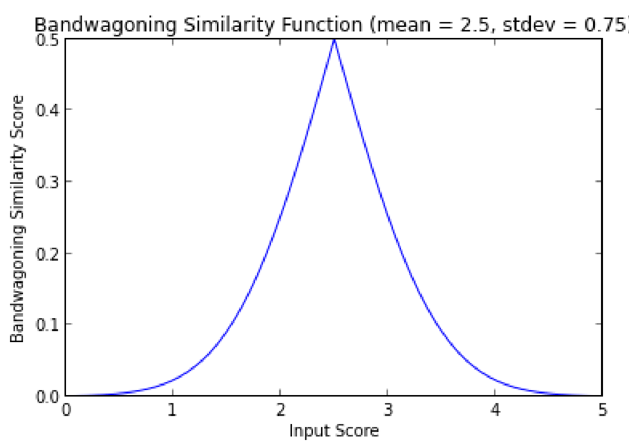


Figure 1: A plot of the bandwagon similarity function for $\mu = 2.5$, $\sigma = 0.75$, as a function of score.

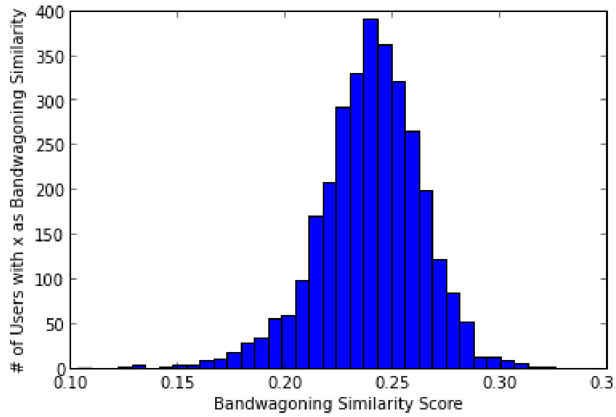


Figure 2: Distribution of user's average bandwagon scores.

2 Prior Work

2.1 Network Analysis

In [1], many natural bipartite networks (e.g. author-paper, actor-movie) and artificial bipartite networks, including those induced by unipartite networks, were studied. It was found that the all bipartite networks, even ones induced by the connections in unipartite networks, had surprising similarities. The authors of this paper also present a model for generating bipartite networks based on the preferential attachment model, which we use for our analysis.

2.2 Online product review environments

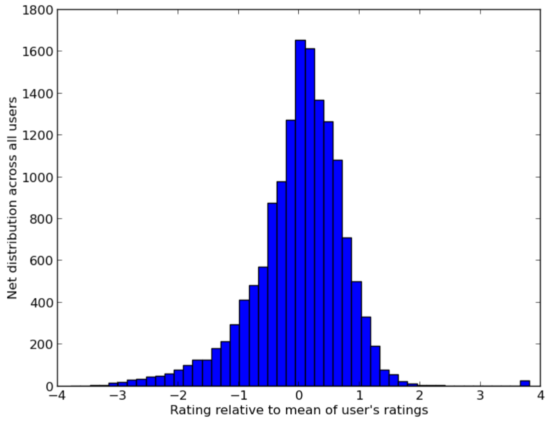
The authors of [3] performed a study on the product reviews written at the website of an online retailer, with a goal of modelling both the incidence (whether a user decides to review) and the evaluation (the actual opinion expressed in the review) of a review. They posit a model where the rating score for a user u on a product b at a time t is dependent on several factors:

- a user bias γ_u
- a set of features $x_{b,t} \in \mathbb{R}^d$ that describes the ratings environment of a user at the time of evaluation. Features used for this include average rating at time of evaluation, standard deviation in rating at time of evaluation, number of items rated at time of evaluation, along with second order combinations of these features.

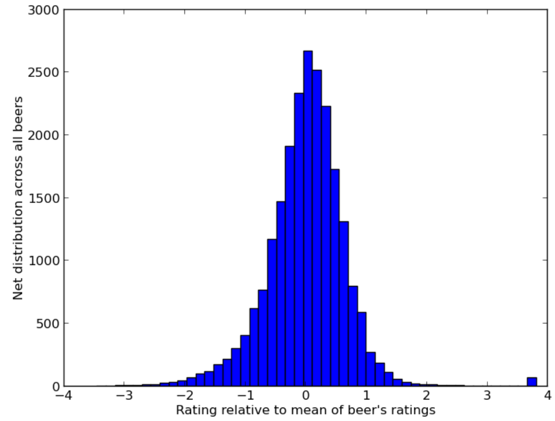
Their main results indicated that less experienced users (where experience is defined according to their model) tended to exhibit bandwagon behavior more often, while more experienced users tended to exhibit differentiation behavior more often. In addition, they provide simulations which indicate that products which have more polarizing ratings tended to be more negative and have a stronger downward trend than compared to a homogeneous neutral customer base.

The results from Moe et al suggest that, if we do observe bandwagon behavior, it is most likely to be among users who have not reviewed many beers, and users who have reviewed more beers should have relatively low bandwagon coefficient. We will analyze our results in terms of Moe et al's results.

3 Initial Data Exploration



(a) Distribution of ratings by user (with >5 reviews)



(b) Distribution of ratings by beer (with >5 reviews)

Figure 3: Basic distribution of rating scores. Note that the distribution of ratings by both user and by beer is normal.

As a first step, we plotted histograms of the ratings of beers relative to mean review score both on a per-user and a per-beer basis. The results are shown in Figure 3. We can determine that in both cases, review scores are distributed approximately normally relative to the mean. Approximately normally distributed review scores are critical for our algorithmic approach, as described when we introduced the bandwagon coefficient.

We also found out that the degree distributions of the users and beers each follow a power-law distribution.

Another interesting property of the beer data is illustrated in Figure 4. The average score of each beer is plotted against the number of reviews written for that beer. We see a positive relation between high number of reviews and average score. Beers that have many reviews are rated highly, whereas beers without many reviews may have either high or low scores. This property supports the existence of bandwagon users who seek out highly rated beers and then review those beers with high scores.

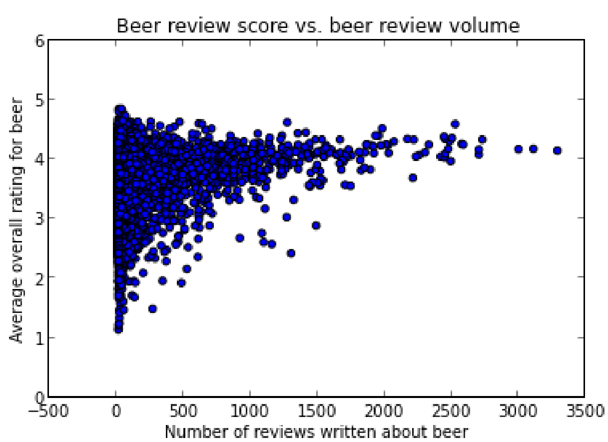


Figure 4: Average rating for each beer vs number of reviews for the beer.

4 Hipsters

The hipster coefficient $h_l(u)$ for a user u has an intuitive definition - the fraction of beers reviewed by u that were reviewed by less than l users.

Figure 5a plots the distribution of hipster coefficient across all users, with $l = 10$. Our choice of $l = 10$ is arbitrary, but was inspired by the `BeerAdvocate` website - only beers with more than 0 reviews are considered for the top 100 ranking on their website. We see a sharp decay with increasing hipster coefficient, with hipster coefficient greater than 0.1 being quite rare.

In order to assess the usefulness of our defined coefficient, we generated bipartite random graphs from several different distributions with comparable numbers of edges and nodes, and compared the resulting distributions of the hipster coefficient across the user nodes. The distributions we compared were:

- An adaptation of the Erdős-Rényi G_{nm} model to bipartite graphs, called the G_{nmk} model, where the two sets N and M in the bipartition have n and m nodes, and k edges with one endpoint in N and the other endpoint in M chosen uniformly at random.

- An adaptation of configuration models for bipartite graphs, where the degree distributions for N and M are already specified, and the nodes in N are wired to those in M uniformly at random.

- An adaptation of the preferential attachment model, where the degree distribution for N (in our case, the users) is already specified, and connections to nodes M are picked, with a connection probability proportional to the current node degree in M.

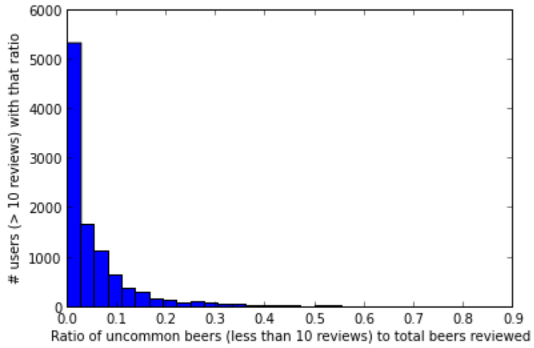
The hipster coefficient distribution for the G_{nmk} model is explained by the fact that most beer nodes generated by this model have degree over 10, making the overall proportion of beers with degree less than 10 comparatively low. The configuration model has the distribution that most visually resembles that of the real-world dataset, although the mode of the distribution is farther from zero. This is explained in part by the fact that the user-beer connections were wired uniformly at random. The fact that the mode of the distribution for the configuration model is larger than that for the `BeerAdvocate` data is consistent with our hypothesis that bandwagon behavior exists in beer reviews - a user exhibiting bandwagon behavior would be less likely to try beers that have not been reviewed as much.

The hipster coefficient distribution for the preferential attachment model also has mean close to zero, but has a second (much less common) mode near 0.2. While it was thought that this model would mimic the real-world model in this respect, in fact this model greatly exaggerates any bandwagon effect that there may be. The explanation for this is that the preferential attachment model results in a smaller proportion of beers that have less than 10 reviews, causing a distribution with a lower and steeper mode.

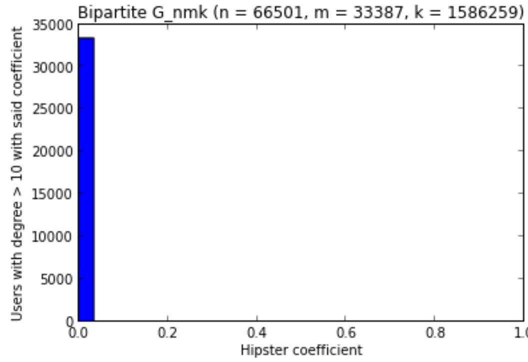
5 Bandwagon Detection

5.1 Model

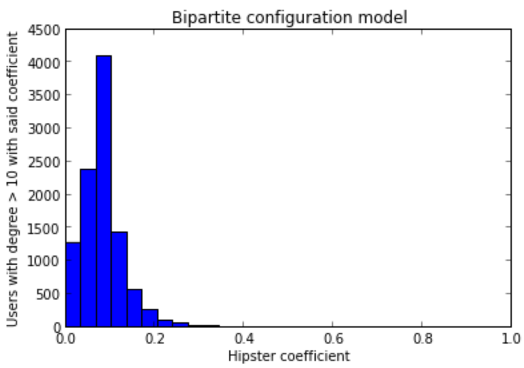
Note that our bandwagon score can be noisy, as it is easy to have cases where a user is honestly reviewing close to the mean for a beer. Thus, we average over a large set of user reviews. Using



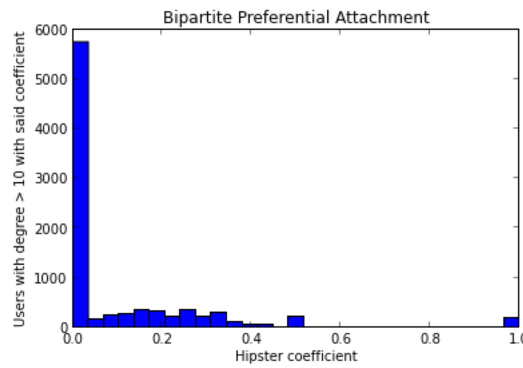
(a) `BeerAdvocate`



(b) Bipartite G_{nmk}



(c) Bipartite Configuration Model



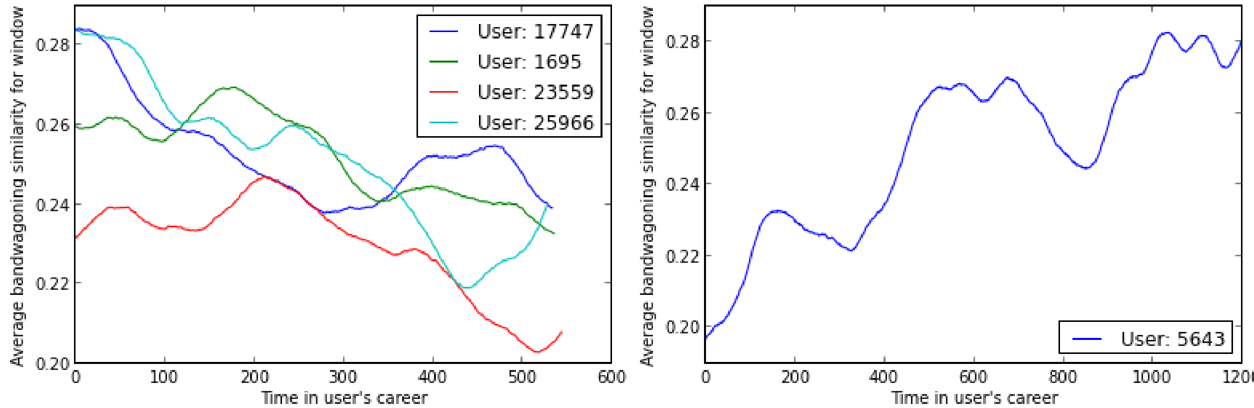
(d) Bipartite Preferential Attachment

Figure 5: Plots of the distribution of the hipster coefficient on natural and artifically generated bipartite networks. (a) is the `BeerAdvocate` dataset. (b) is a bipartite version of the G_{nmk} model, with $n = 66501$ beers, $m = 33387$ users, and $k = 1586259$ edges between users and beers. (c) is a graph that was randomly wired with degree distributions for users and beers already given. (d) is the bipartite preferential attachment model described in [1] with the user degree distribution already specified.

this similarity score, one of the methods we use to observe bandwagoning behavior is to examine the evolution the bandwagoning similarity score a user receives over time. This approach allows us to observe whether users who exhibit bandwagoning behavior are bandwagoners for their entire careers on `BeerAdvocate` or if users become more deviant from mean reviews as they write more reviews.

5.2 Experiments

For our first bandwagoning experiment, we considered only users who have reviewed more than 400 beers. A plot for this experiment is shown in Figure 6a. We convolve the users reviews with a Tukey window and observe the trends of their average bandwagoning similarity as as their beer reviewing careers progress. This method reduces noise of the similarity metric while also allowing us to generalize the bandwagoning of a user for a window of time. Many of the users show general downward trend. This shows that when these users first joined `BeerAdvocate`, their review scores were very similar to the average of that beer, but they began to differentiate over time. It is also worth noting that the bandwagoning similarity can be compared among users, e.g. user 23559



(a) Visualizations of user bandwagon coefficient for four users as they write more reviews.

(b) Visualizations of user 5643's bandwagon coefficient they write more reviews. Note the upward trend.

Figure 6

always displayed more differentiating behavior than user 1695.

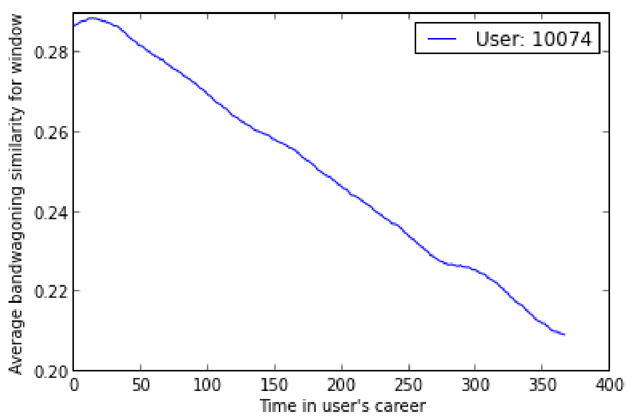


Figure 7: Visualizations of user 10074's bandwagon coefficient they write more reviews. Note the strictly downward trend.

However, the plots in Figure 6a do not encompass the trends for all users as they write more reviews. Figure 6b shows the same plot for user 5643, who exhibits an upward trend for bandwagon score over time. We hypothesize that user 5643 began rating on `BeerAdvocate` as a novice and gradually learned to reviews properly, converging on more similar ratings to other users over time. Figure 7 shows a similar plot for user 10074, where we see an almost strictly decreasing bandwagon coefficient over time. Thus, some users exhibit starting off as bandwagoners and show more differentiation over time, while some others show the opposite effect.

Finally, we visualize hipster coefficient vs. bandwagon score in Figure 8 to test our hypothesis that users with low bandwagon score tend to have high hipster coefficient. This experiment was computed for users who have reviewed at least 100 beers. This experiment includes $\sim 85\%$ of all `BeerAdvocate` reviews. We notice a distinct negative relationship between hipster coefficient and

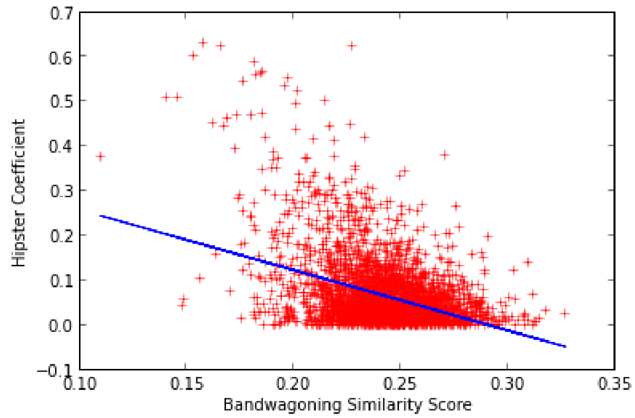


Figure 8: Scatter plot of hipster coefficient vs bandwagon score for users authoring more than 100 reviews, with a linear regression of linear fit.

bandwagon score. In particular, a low density of "beer hipsters" with high hipster coefficient have the some of the lowest bandwagon scores.

6 Conclusion

In our study, we set out to determine if there was a relationship between the mindset of how a user selects the beers he tries (modeled in this instance by our hipster coefficient) and the mindset of how that user determines the score of that review (modeled in this instance by the bandwagoning similarity score). We analyzed our choices in metrics and demonstrated why they were reasonable as well as how they look within the beer dataset. Finally, we compared the values of these metrics for each user with a substantial career and showed that a relationship does exist between the hipster coefficient and bandwagoning similarity of a user, where a higher bandwagoning similarity makes a lower hipster coefficient more probable.

The implication here is a correlation, but there are many intuitive explanations that may have caused it. Where Moe and Schweidel [3]. focused on linking the evaluation of a product with the incidence of the review given that the product has been purchased and evaluated, we believe our relationship adds to this by linking the evaluation of a product with the decision to purchase it (as in the choice to review a beer with few reviews as opposed to one with many). We believe that this relationship shows that, in the same way a user may generate his review based on prior reviews, a users mindset in selecting what beer to review may also be influenced by the ratings of the dataset. In other words, a review system exists in an open feedback loop with the set of users as opposed to being a true static model of how an isolated user would feel about an isolated beer. In this case, we see that users who are less likely to care about the current reviews of a beer when rating a beer are also the users who are less likely to base their selection in beer on the current set of reviews. Another reasonable implication is that these users are purposefully differentiating themselves from the mean and purposefully selecting beers with low reviews to try to add diversity to the data set and/or differentiate themselves from other users.

References

[1] J. Guillaume and M. Latapy. Bipartite structure of all complex networks. *Information processing letters*, 90(5):215–221, 2004.

[2] J. McAuley, J. Leskovec, and D. Jurafsky. Learning attitudes and attributes from multi-aspect reviews. *arXiv preprint arXiv:1210.3926*, 2012.

[3] W. Moe and D. Schweidel. Online product opinions: Incidence, evaluation, and evolution. *Marketing Science*, 31(3):372–386, 2012.