Predicting Consumer Sentiment Towards Amazon Fashion Products Using a Product-Reviewer Network

Zoey Li[a], **Vincent Chiu**[b], **and Hinal Jajal**[c]

[a,b,c]Department of Mathematics, University of California, Los Angeles, CA 90024

This manuscript was compiled on December 16, 2022

Product reviews form a goldmine of text data that is helpful to both consumers and businesses in making buying decisions and understanding consumer needs respectively. In this study, we present a method to convert review text data into a product-reviewer network that can be used to predict consumer sentiment towards a product they have not previously reviewed. Our method utilizes two key techniques in this process: the first is natural language processing to extract the sentiment from thousands of reviews and the second is balance theory to analyze the signed network. We address the challenges in drawing insights from a signed, bipartite network and make use of existing literature on modified balance theory to overcome them. With a carefully sampled dataset of Amazon fashion products' reviews, we demonstrate the effectiveness of our method in predicting the signs of consumer sentiment. We explore three sign prediction methods, namely the Signed Caterpillar (SCsc), Random Walk Based Models (SBRW), and Matrix Factorization (MFwBT), and successfully implement two of the three with remarkable accuracy. We recognize the several limitations in our processes, including discrete sentiment values and a limited dataset. Future work would expand these promising results into a more comprehensive study addressing the mentioned limitations.

Sign Prediction | Undirected Signed Bipartite Network | Balance Theory | Signed Butterfly | Signed Caterpillars |Sentiment Analysis | Amazon Fashion

Text data has become increasingly available and relevant in the past decade with the rise of global internet. Social media posts, review websites, and online surveys have all contributed to the wealth of information we can use to understand evolving human behavior. Of particular interest to businesses have been reviews and ratings, which aid in constructing recommender systems (2), increasing sales (3), and improving consumer satisfaction. Reviews — rather than ratings — are especially useful to analyze because words provide a more meaningful interpretation that is masked by one-dimensional ratings. There is an abundance of research on sentiment analysis of product reviews which has achieved remarkable results using state-of-the-art machine learning techniques (4), deep neural networks (5), and feature specific analysis (6). However, existing work has been lacking in either one of two aspects. On one hand, previous studies have exclusively focused on accurately classifying reviews as positive or negative rather than predicting the sentiment of consumers towards unfamiliar products. On the other hand, businesses' internal systems that do prioritize the latter exploit consumer metadata such as demographics (2). We contribute to the existing research in a multi-fold manner. Firstly, we create a process to predict consumer sentiment towards unknown products. Secondly, we achieve this by exploiting the network structures of products and reviewers rather than their metadata. Previous research on the ethics of collecting and using user data has revealed the existence of risks related to privacy and opacity (7). By relying solely on the review text and no identifying information, we hope to mitigate such risks in our method while providing an accurate, powerful predictive functionality. Another motivation to make use of network science in drawing from the results of sentiment analysis includes the proven wide-ranging benefits of the analysis of social networks and technological networks (1, 10). Previous network studies in e-commerce have also understood online reviewer characteristics which would not have been possible without a network position analysis (8).

In our attempts to take advantage of the network structures, we encounter various challenges owed to the negative weights and bipartite sets in our constructed network. Recent work of Derr et. al in extending the notions of balance theory to bipartite networks using signed butterflies instead of signed triads proved especially helpful. We draw on the three sign prediction methods created by them: the Signed Caterpillars Based Classifier, Random Walk Based Signed Prediction, and Matrix Factorization with Balance Theory. They tested and

Significance Statement

The significance of our study can be understood from two points of view. From a theoretical standpoint, existing work on signed, bipartite networks is limited. Expanding existing network functionalities for such a network is challenging with the presence of negative edge weights and disjoint sets of nodes. At the same time, applications of such a network are ubiquitous — from terrorist-target to buyer-seller networks — making the learning of its construction and analysis important. In this case, we focus on the application of analyzing signed, bipartite networks in e-commerce. We make use of advanced sign prediction methods and extend their application to a product-reviewer network. From an application standpoint then, we better understand consumer preferences and aid in decision making while circumventing the need for any user data.

Z.L., V.C., and H.J. designed research, performed research, analyzed data, and wrote the paper. Z.L did the data visualization, implemented the SCsc method, co-wrote the Abstract, Introduction, and Background information, and wrote the Data Visualization, SCsc methodology, and Limitation sections in this report; V.C. did the sentiment analysis and data preprocessing, implemented the MFwBT method, wrote the Data Acquisition, Sentiment Analysis, MFwBT Methodology, Results, and Acknowledgements sections in the report; H.J. acquired the data, assisted with the data preprocessing and sentiment analysis, implemented the SBRW method, performed the signed butterfly calculations, co-wrote the Abstract, Introduction, Background information sections, and wrote the Balance Theory, SBRW Methodology, and Conclusion sections in the report

The authors declare no conflict of interest.

[1]Zoey Li (Author One) contributed equally to this work with Vincent Chiu (Author Two) and Hinal Jajal (Author Three).

[2]To whom correspondence should be addressed. E-mail: zoeyli@g.ucla.edu

applied these methods to three datasets. One of these is the Bonanza dataset belonging to a shopping website similar to Amazon, while the other two differ in that they are datasets covering vote information for political purposes (9). As described in the work of Derr et al, these methods that receive aid from balance theory perform better than their respective baseline methods (9). The good performance of these models, especially on the Bonanza dataset, motivates our choice to apply them in the Amazon setting.

Background Information

We limit the scope of our analysis to the Amazon Fashion Products dataset (13). Specifically, we select a 5-core, randomly sampled reviews dataset. The small size of the subset enables our processes to be computationally feasible and its density (k-core) provides repeated measures that are crucial in making consumer-specific predictions.

Prior to creating the network, we perform sentiment analysis, which uses natural language processing to determine whether the sentiment of a text is positive or negative. Based on the results, we create an undirected signed bipartite network. In the construction of our network, we draw inspiration from the work of Wang on Yelp reviews (8). Similarities include incorporating a bipartite network with two distinct types of nodes. The first group of nodes would be the reviewers and the second group of nodes would be the Amazon fashion products. An edge represents that a reviewer wrote review for that specific product. The network is undirected because this relationship is nondirectional by nature. Our work varies from Wang's in the assignment of edge weights due to the additional component of reviewer sentiment. The network is weighted, and the weights are the sentiment values derived from the results of sentiment analysis. We restrict the edge weight to be either -1 or -1, with 1 representing positive sentiment and -1 representing negative sentiment. It would be more reliable to have values between -1 and 1 in order to emphasize different degrees of sentiment toward the products. However, we opt for two discrete values to simplify the sign prediction process.

At the beginning, after constructing such a network, we aimed to utilize centrality measures to rank products based on reviewer generosity, and identify reviewers with the most extreme sentiments. However, since our network involves signs and is bipartite, there's no suitable centrality measures available for us to use. Thus, we modify our topic to make sign predictions which is equally interesting and has existing literature.

Materials and Methods

Data Acquisition. The dataset used in this project comes from Dr. Julian McAuley's research group at UC San Diego (12). Professor McAuley is a computer scientist whose research focuses on machine learning and recommender systems. His website contains a wide variety of datasets spanning from Google local reviews to Food recipes. To be allowed to use his dataset, he has requested that his works regarding the dataset to also be cited. His dataset regarding Amazon product reviews contains over 80 million reviews from over 20 million users. Professor McAuley et al. has used this data to study the evolution of fashion trends using one-class collaborative filtering and image-based recommendations (14, 15).

The dataset we use is a smaller subset of only 3000 product reviews coming from the fashion category (13). The dataset contains information such as the style and size of clothing, review date, ratings, etc., but the only information we need to construct our bipartite network is the product, reviewer, and the review sentiment.

A. Sentiment Analysis. As the product names and reviewers are already in the dataset, we still need to find the review sentiment for each product review. This is accomplished by extracting positive or negative sentiment from the review text, also known as sentiment analysis. Sentiment analysis uses natural language processing to determine if the sentiment of a piece of text is positive or negative. Some of the most common sentiment analysis libraries in Python include nltk, textblob, bert, etc. This project constructed sentiment bipartite networks using nltk.

To construct the bipartite network using nltk, we treat the sentiment analysis as a classification task where we build a machine learning model to predict whether a given product review is positive or negative (17). As sentiment analysis is a supervised method, we require target variables. For our case, the target variables that the machine learning model will be trained on will be based on the ratings of the product review. Because most of the ratings rated the products as 4/5 or 5/5 stars, we decide to categorize ratings of 3 stars or below as negative reviews. This decision is made in hopes that this will create a more balanced network instead of one that is overwhelmingly positive. A possible reason why product reviews are generally positive could be that buyers who are unsatisfied with their purchase simply requested a return and never left a negative feedback.

Next, we generated the term document matrix using count vectorizer in the scikit learn library. Term document matrix is a data frame that counts the occurrences of words across all texts and outputs a matrix where the rows are the text it came from and the columns are the words that appeared in the text. By our construction, we disregard words that appear rarely as well as words that appear too frequently. We also ignore English "stopwords", which are words that are frequently use but don't provide any contextual information such as "the, I, and, too, etc." This results in a term document matrix as seem below. The matrix contains over 3000 reviews and over 200 predictor words.

	10	able	absolutely	actually	amazing	amazon	ankle	arch	arches	area	...	wish	wore	work	worked	working	workout	workouts	worn	years
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3171	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
3172	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
3173	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
3174	0	0	0	0	0	0	1	0	0	0	...	0	0	0	0	0	0	0	1	0
3175	0	0	0	0	0	0	0	0	0	0	...	0	0	1	0	0	0	0	0	0

Fig. 1. Term Document Matrix

We select logistic regression as the machine learning model for sentiment analysis. The reason behind choosing logistic regression and not support vector machine or neural networks is because we are able to easily interpret the sentiment of individual words using the coefficients of the logistic model. These coefficients are the weights of each individual word used in model prediction to determine whether the entire review is positive or negative. Tables 1 and 2 show the top 5 most positive and top 5 most negative words found after conducting sentiment analysis.

Table 1. Words in Term Document Matrix with Most Positive Sentiments

coef	word
2.249833	perfect
2.014510	lightweight
1.667149	smaller
1.655638	make
1.617396	white

Table 2. Words in Term Document Matrix with Most Negative Sentiments

coef	word
-2.969220	returned
-2.750874	left
-2.483192	fine
-2.061016	support
-2.008989	large

Notice that while some sentiments determined by the model intuitively make sense, such as "perfect" and "returned" being associated with positive and negative reviews, there are also words whose sentiments aren't very obvious such as "smaller" being positive and "large" being negative. After tuning the complexity and doing 5-fold cross validation, the model is able to achieve a 96.8% accuracy in identifying positive and negative product reviews. We assign the positive reviews with edge weight of 1 and negative reviews with edge weight of -1, combined with the product and reviewer information, we have constructed the bipartite signed network.

We also conducted sentiment analysis using the textblob library in Python, where we input the review text and the model outputs a polarity score. This however did not provide as much insight as the model we built using nltk, so it was not used in later steps.

Data Visualization. The Figure 2 is generated by using Python and NetworkX. It has 368 nodes and 2484 number of edges in total, so the edges are densely distributed in the network. The left group of nodes are the Amazon Fashion Products, and the right group of nodes are the reviewers. Since the network is bipartite, there could only be edges between different types of nodes.

The network is composed of only two colors, red and green. An red edge means that the reviewer wrote a negative review for that product, and a green edge means that the reviewer wrote a positive review for that product. In this network, it can be clearly seen that most of the edges are green with only a few red edges, which means that people prefer to give positive feedback in general for Amazon Fashion Products. This is also true in reality where we can see most people tend to give good comments when writing reviews for the products.

Also, we can see the green tend to be deeper in the left group of nodes because these nodes representing Amazon Fashion Products have much higher degrees than the reviewers, which means there are much more edges attached to them.

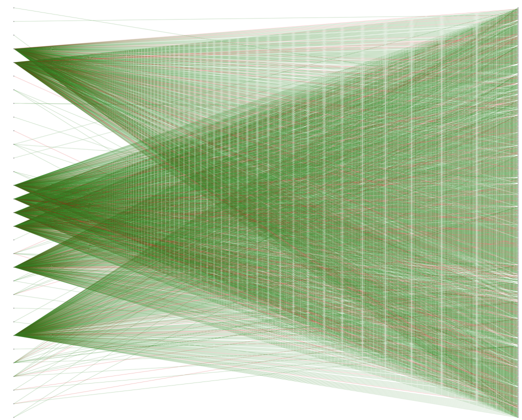

Fig. 2. Undirected Signed Bipartite Network Visualization

The Figure 3 is the degree histogram. It is drawn by using Python and NetworkX. We can see that most of the nodes have degree less than 20, and only a very few have degree around 300, so the histogram becomes loosely distributed.

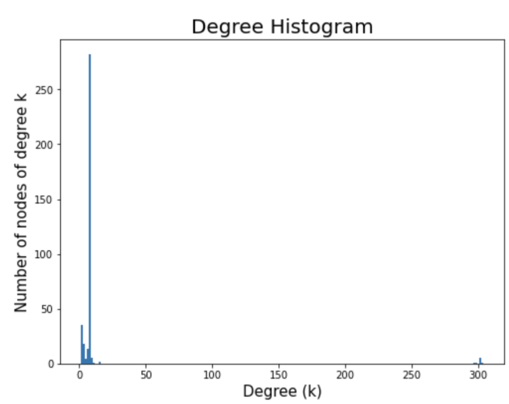

Fig. 3. Degree Histogram

The Figure 4 is the log-log degree distribution plot. Most of the nodes in this network have degree less than 20. However, some of the nodes have degree around 300 that are Amazon Fashion Products which receive high amount of reviews from customers. Thus, there can be a really large gap between these degree values, so it would be more reasonable to use a log-log plot instead of the original one.

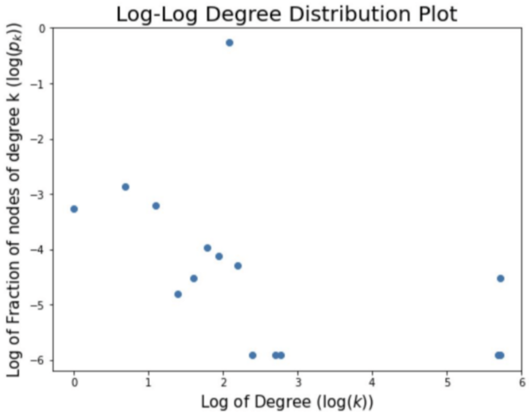

Fig. 4. Log-Log Degree Distribution Plot

Balance Theory. The study of signed networks is challenging because we lose access to otherwise useful tools such as centrality measures. We utilize balance theory because it is designed for signed networks and is the most advanced technique in their study (11). Balance theory uses the notion of tension in social systems to classify these systems as either balanced – which are less likely to change – or unbalanced. Applying lessons from traditional balance theory to bipartite networks, however, is not straightforward. Bipartite networks have two key hindrances in that they have two different types of nodes and do not have triads. Derr et. al expand the functionality of balance theory to bipartite networks with the creation of signed butterflies as illustrated in Fig. 5. In 5(a), the structure with all positive signs represents that both the reviewers reviewed the two products positively. Unbalanced structures in 5(f) and 5(g) represent unstable scenarios where two reviewers possess opposing views on the same products. In real-life, we expect unbalanced structures to occur less frequently due to their unstable nature (1). We count the instances of different classes of balanced and unbalanced signed butterflies in our dataset. Table 3 verifies that for the Amazon fashion products dataset, we observe a significantly greater proportion of balanced scenarios. We show that the dataset adheres to balance theory defined in terms of signed butterflies, making balance theory based sign prediction methods viable in this context.

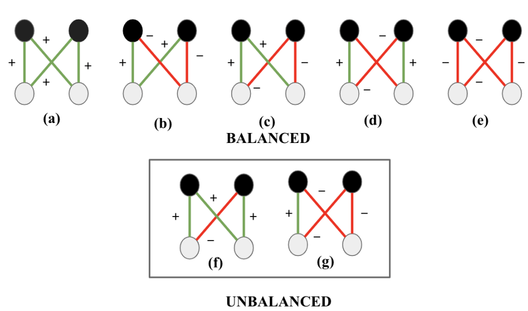

Fig. 5. The Seven Signed Butterfly Structures

Sign Prediction Methods.

Table 3. Signed Butterfly Distribution in the Amazon Products Dataset

Signed Butterfly Class	Count	%
(a) (+,+,+,+)	3,573,088	83.3
(b) (+,-,-,+)	609,556	14.2
(c) (+,+,-,-)	16	3.73e-04
(d) (+,-,+,-)	0	0
(e) (-,-,-,-)	101,196	2.34
Balanced	**4,283,856**	**85.6**
(f) (+,+,+,-)	4072	0.09
(g) (+,-,-,-)	682	0.02
Unbalanced	**4,754**	**0.11**

A.1. Signed Caterpillars Based Classifier Method. The first sign prediction method that will be covered in this project is the SCsc method, which is short for Sign Caterpillars Based Classifier. A signed caterpillar is paths of length 3 that are missing just one link to become a signed butterfly that's covered above in balance theory, and it can either have balance path or unbalanced path (9). A balanced path means presence of even number of negative links among those three, while an unbalanced path means presence of odd number of negative links among those three. In this method, the signs are predicted by extracting features from either the individuals (i.e. their positive or negative degrees) or local neighborhood features based on balance theory (i.e. signed caterpillars), and Logistic regression machine learning model is built to make final predictions (9). This method is successfully implemented by modifying the existing coding template provided in the work of Derr et al (16). We applied this method by first dividing the original Amazon Fashion dataset into two different sets, with 90 percent as the training set and the rest 10 percent as the test set, so we can check the performance of our model based on the test set.

A.2. Random Walk Based Signed Prediction Method. Random-walk-based methods are a ubiquitous solution for various purposes including link prediction. Previous applications of these methods have been limited to link prediction in unsigned unipartite networks. Derr et al. extend their application to sign prediction and incorporate balance theory (9). We reproduce their method for the Amazon dataset and detail the corresponding process below.

To make use of the random-walk-based models, the bipartite network $\mathbf{B}$ is first converted into a unipartite one $\mathbf{A}$. This is achieved by creating one-mode projection matrices for $\mathbf{U}_P$ and $\mathbf{U}_R$, which are the sets of product and reviewer nodes respectively. Then, the projection matrices are converted into an adjacency matrix $\mathbf{A}$. To create the one-mode projection matrices, Derr et al. use balance theory to form signed triangles. Let ns^A_{ij} be the number of products that reviewers i and j agree on. Conversely, let ns^D_{ij} be the number of products they disagree on. Then $\mathbf{P}_{Rij} = \mathbf{P}_{Rji} = ns^A_{ij} - ns^D_{ij}$ where $\mathbf{P}_R$ is the reviewer projection matrix. In words, $\mathbf{P}_R$ represents the degree of agreement between different reviewers. We can construct a products projection matrix $\mathbf{P}_P$ in a similar manner. The real-life interpretation of such a matrix is not as straightforward as that of reviewers, but it can be understood to represent the degree of similarity in being liked or disliked between different products. To avoid adding trivial connections, we bound $ns^A_{ij} - ns^D_{ij}$ using the following definition:

$$\mathbf{P}_{Rij} = \begin{cases} 0 & \text{if } \delta_{low} < ns^A_{ij} - ns^D_{ij} < \delta_{high} \\ ns^A_{ij} - ns^D_{ij} & \text{otherwise} \end{cases} \quad [1]$$

where we arbitrarily pick $\delta_{low} = -10$ and $\delta_{high} = 10$.

The projection matrices $\mathbf{P}_R$ and $\mathbf{P}_P$ are converted into adjacency matrix $\mathbf{A}$ according to the following definition:

$$\mathbf{P}_{Rij} = \begin{bmatrix} \hat{\mathbf{P}}_B & \omega\hat{\mathbf{B}} \\ \omega\hat{\mathbf{B}}^T & \hat{\mathbf{P}}_S \end{bmatrix} \quad [2]$$

where $\hat{\mathbf{B}}$ is the row normalized version of $\mathbf{B}$, defined as $\hat{\mathbf{B}}_{ij} = \mathbf{B}_{ij} / \sum_k |\mathbf{B}_{ik}|$ and ω is a parameter created to favor the real, existing links rather than those inferred using balance theory. Finally, we use a random-walk-based model and define $\mathbf{Y}$:

$$\mathbf{Y}_{ij} = \sum_k \hat{\mathbf{A}}_{ik}\mathbf{Y}_{kj} \quad [3]$$

We obtain the link sign predictions from the upper right corner of $\mathbf{Y}$.

A.3. Matrix Factorization with Balance Theory Method. Matrix Factorization with Balance Theory, or MFwBT, is a method that expands on the Basic Matrix Factorization model by incorporating Balance Theory (9).

The matrix factorization model approach considers the following optimization problem:

$$\min_{U,V} \sum_{(p_i,r_j)\in\mathcal{E}} \max\left(0, 1 - \mathbf{B}_{ij}(\mathbf{u}_i^\top \mathbf{v}_j)\right)^2 + \lambda(|\mathbf{U}|_F^2 + |\mathbf{V}|_F^2) \quad [4]$$

Where the objective is to discover the latent matrices of the set of products and reviewers, $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_{n_P}] \in \mathbb{R}^{d\times n_P}$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{n_R}] \in \mathbb{R}^{d\times n_R}$ for dimension d. Given $\mathbf{u}_i^\top \mathbf{v}_j$ is the predicted link sign between product p_i and reviewer r_j, and $\mathbf{B}_{ij}$ is the real link sign between product p_i and reviewer r_j, summed over $\mathcal{E}$ the set of edges in the the bi-adjacency matrix $\mathbf{B}$. If the predicted link and the real link are of the same sign and greater than 1, there is no loss. If the predicted link and the real link have different signs, then the loss value will drive the minimization during the training process. This minimization can be achieved using Stochastic Gradient Descent (SGD) following the works of (18).

The shortfall of matrix factorization is that it focuses on minimizing the errors of predicting existing link signs, by incorporating balance theory, it can convert many signed caterpillars into balanced signed butterflies to encourage the learning of link signs of product and reviewer pairs that previously did not exist.

The Matrix Factorization with Balance Theory method is given by:

$$\begin{aligned} &\min_{U,V} \sum_{(p_i,r_j)\in\mathcal{E}} \max\left(0, 1 - \mathbf{B}_{ij}(\mathbf{u}_i^\top \mathbf{v}_j)\right)^2 + \lambda(|\mathbf{U}|_F^2 + |\mathbf{V}|_F^2) \\ &+ \alpha \sum_{(p_i,r_j)\in\hat{\mathcal{E}}_i^+} \max\left(0, 1 - \hat{\mathbf{S}}_{ij}(\mathbf{u}_i^\top \mathbf{v}_j)\right)^2 \\ &+ \beta \sum_{(p_i,r_j)\in\hat{\mathcal{E}}_i^-} \max\left(0, 1 - \hat{\mathbf{S}}_{ij}(\mathbf{u}_i^\top \mathbf{v}_j)\right)^2 \end{aligned} \quad [5]$$

Where α, β are weights used to control the incorporation of signed butterflies $\hat{\mathbf{S}}$ using implicit positive and negative links $\mathcal{E}_i^+$ and $\mathcal{E}_i^-$ as defined by the balance theory.

The implementation of MFwBT ran into trouble because the code implementation provided by Derr et al. requires an "extra_links_from_B_balance_theory.txt" file (16). Based on inference, we presume these extra link files are the implicit positive and negative links $\mathcal{E}_i^+$ and $\mathcal{E}_i^-$. However, the authors did not explain how these extra links are generated and there is no example of how such a link file would be formatted. Had I, Vincent, noticed this missing detail earlier, we might have had enough time to generate some form of extra link file through pure trial and error to get some result. As a result, we failed to implement this method with our signed bipartite network.

Results

The results are given by two metrics, the AUC score and F1 score. AUC, or the "Area Under the ROC Curve," measures the two-dimensional area under the ROC (receiver operating characteristic) curve. The ROC curve plots True Positive Rate (TPR) vs. False Positive Rate (FPR) on a scale from 0 to 1 for different classification thresholds (19). True Positive Rate and False Positive Rate are given by the formulas:

$$\begin{aligned} \text{TPR} &= \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \\ \text{FPR} &= \frac{\text{False Positive}}{\text{False Positive} + \text{True Negative}} \end{aligned} \quad [6]$$

In simplest terms, if AUC is 0.0, the classifier is misclassifying all positives as negatives, and all negatives as positives; if AUC is 0.5, the classifier is classifying values at a rate no better than a coin flip; if AUC is greater than 0.5, the classifier is classifying most values correctly, and if AUC is exactly 1, it is classifying all values correctly (21).

The other metric is the F1 score, which is the harmonic mean of precision (P) and recall (R) (20):

$$\frac{1}{F_1} = \frac{\frac{1}{P} + \frac{1}{R}}{2} \quad [7]$$

Note precision is the percentage of positively classified that are actually positive:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad [8]$$

And recall is the percentage of actual positives that are correctly classified:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad [9]$$

F1 score is useful for imbalanced data because it is a better metric than plain accuracy.

The result of the code implementation of our sign prediction method is given by Table 4.

Observe that the Signed Bipartite Random Walk Method (SBRW) performed better than the Signed Caterpillar based Classifier Method (SCsc) on both metrics. This is different from the result of the paper where the methods are based on (9), where no one single method outperformed others across all datasets and metrics. However, considering that we only

Table 4. Link Sign Prediction Results in terms of (AUC, F1)

Metric	SCsc	SBRW	MFwBT
AUC	0.864	0.970	-
F1	0.877	0.993	-

have one dataset and implemented two of the three methods, it is probably not out of the ordinary.

Also notice that F1 score tends to be higher than AUC score. This confirms the results from Derr et al.'s paper where the F1 scores are much higher than AUC in the Bonanza dataset (9). In the paper, the Bonanza dataset has a heavy positive imbalance where 98% of the links are positive. This is similar to our dataset where 85% of the links are positive. In cases like these, the AUC is a more accurate metric than F1. We can see that to better detect negative links, some positive links are misclassified.

According to the literature, the Matrix Factorization with Balance Theory method (MFwBT) is supposedly better at balancing the ratio of positive and negative implicit links depending on the choice of α and β (9). However, this is hard to confirm because we did not successfully implement the method and the paper itself fixed $\alpha = \beta$ for their experimentations.

Conclusion

In conclusion, the outcomes of this paper are illuminating on multiple fronts. Firstly, we devise a method to study reviewer sentiments in a way that utilizes network features. Specifically, we construct an undirected, signed, bipartite product-reviewer network where the signs represent discrete values of reviewer sentiment. We also carefully extract sentiment values using natural language processing. Secondly, we conduct exploratory analysis of the network, generating insights on the degree distribution, sentiment (sign) distribution, and adherence to balance theory. Finally, we explore three sign prediction methods and successfully implement two with more than 85% accuracy. The accuracy and effectiveness of Amazon's prediction systems likely remain confidential, making comparison of our models with theirs difficult. We are also not aware of any other academic work in predicting consumer sentiment that our results can be compared with. Nonetheless, with a limited number of features and an avoided risk of privacy issues, our models obtain an objectively reliable accuracy. The code materials and data to reproduce our results can be found at this link.

Limitations

The first limitation of our project is that the edges weights of our undirected signed bipartite network are constrained to be either 1 or -1. In reality, there should be values between -1 and 1 to represent different levels of positive sentiment and negative sentiment. In our case, a reviewer could extremely like a product or just kind of like the product which make a huge difference. However, we only use 1 and -1 in this project in order to simplify our sign prediction process.

The second limitation is that we only use a small subset for experimentation with only thousands of product reviews. In reality, there would be much larger dataset which cannot be tested in our own computer. The machine learning models would be more accurate if we can apply sign prediction to a larger dataset.

The third limitation is that when a person rates a product multiple times, we only keep the first sentiment value in order to avoid multi-edges.

ACKNOWLEDGMENTS. We acknowledge Julian McAuley for sharing the Amazon Fashion Product Review data that was collected for ref. (13–15). We also thank Derr et al. for sharing their code implementations (9, 16). We thank Mason Porter and Grace Li for helpful comments and advice, generous extension, and the formatting examples (22, 23). This report was made possible by https://www.overleaf.com.

1. M.E.J. Newman, Networks. *Oxford University Press*, 2 ed. (2018).
2. Stratigi, M., Li, X., Stefanidis, K., Zhang, Z. (2019). Ratings vs. Reviews in Recommender Systems: A Case Study on the Amazon Movies Dataset. In: , et al. New Trends in Databases and Information Systems. ADBIS 2019. Communications in Computer and Information Science, vol 1064. Springer, Cham. https://doi.org/10.1007/978-3-030-30278-8_9
3. Sharma, Kanak. Future sales prediction of OnePlus Using Amazon Review's data. Diss. Dublin Business School, 2018.
4. Jagdale, R.S., Shirsat, V.S., Deshmukh, S.N. (2019). Sentiment Analysis on Product Reviews Using Machine Learning Techniques. In: Mallick, P., Balas, V., Bhoi, A., Zobaa, A. (eds) Cognitive Informatics and Soft Computing. Advances in Intelligent Systems and Computing, vol 768. Springer, Singapore. https://doi.org/10.1007/978-981-13-0617-4_61
5. Onan, Aytuğ. "Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks." Concurrency and Computation: Practice and Experience 33.23 (2021): e5909.
6. Mukherjee, S., Bhattacharyya, P. (2012). Feature Specific Sentiment Analysis for Product Reviews. In: Gelbukh, A. (eds) Computational Linguistics and Intelligent Text Processing. CICLing 2012. Lecture Notes in Computer Science, vol 7181. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-28604-9_39
7. Janice C. Sipior, Burke T. Ward & Nicholas M. Rongione (2004) Ethics of Collecting and Using Consumer Internet Data, Information Systems Management, 21:1, 58-66, DOI: 10.1201/1078/43877.21.1.20041201/78986.6
8. Wang HJ. Understanding reviewer characteristics in online reviews via network structural positions. Electron Mark. 2022;32(3):1311-1325. doi: 10.1007/s12525-022-00561-z. Epub 2022 Jun 27. PMID: 35789703; PMCID: PMC9243715.
9. T. Derr, C. Johnson, Y. Chang, J. Tang, Balance in Signed Bipartite Networks. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. pp. 1221-1230 (2019).
10. M. Szell, R. Lambiotte, S. Thurner, Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences.* 107(31): 13636-13641 (2010).
11. S. A. Marvel, J. Kleinberg, R. D. Kleinberg, S. H. Strogatz, Continuous-time model of structural balance. *Proceedings of the National Academy of Sciences.* 108 (5): 1771-1776 (2011).
12. J. McAuley, et al., Recommender Systems and Personalization Datasets. (2018). https://cseweb.ucsd.edu/~jmcauley/datasets.html#amazon_reviews (accessed 1 December 2022).
13. J. Ni, J. Li, J. McAuley, Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP).* pp. 188-197 (2019).
14. R. He, J. McAuley, Ups and Downs: Modeling the Visual Evolution of Fashion Trends with One-Class Collaborative Filtering. *Proceedings of the 25th International Conference on World Wide Web.* pp. 507-517 (2016).
15. J. McAuley, C. Targett, Q. Shi, A. van den Hengel, Image-based Recommendations on Styles and Substitutes. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval.* arXiv:1506.04757 (2015).
16. T Derr, et al., Balance in Signed Bipartite Networks. (2019). https://github.com/DSE-MSU/signed-bipartite-networks (accessed 1 December 2022).
17. P.S. Chodrow, Sentiment Analysis. *UCLA Fall 2021 PIC16A Course Notes* (2021) (accessed 1 December 2022).
18. C.J. Hsieh, K.Y. Chiang, I.S. Dhillon, Low rank modeling of signed networks. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* pp. 507-515 (2012).
19. Google Machine Learning Education, Classification: ROC Curve and AUC. https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc (accessed 10 December 2022).
20. K.W. Chang, et al., Lecture 8: Neural Network & Deep Learning. *UCLA Fall 2022 CM146 Course Notes* pp. 33-36 (2022).
21. A. Bhandari, AUC-ROC Curve in Machine Learning Clearly Explained. https://www.analyticsvidhya.com/blog/2020/06/auc-roc-curve-machine-learning/ (accessed 10 December 2022).
22. E. Berthier, M.A. Porter, K.E. Daniels, Forecasting failure locations in 2-dimensional disordered lattices. *Proceedings of the National Academy of Sciences.* 116(34): 16742-16749 (2019).
23. J.P. Gleeson, D. Cellai, J.P. Onnela, M.A. Porter, F. Reed-Tsochas, A simple generative model of collective online behavior. *Proceedings of the National Academy of Sciences.* 111(29): 10411-10415 (2014).
24. H. Zhang, J. Chen, Bipartite Consensus of Linear Multi-Agent Systems Over Signed Digraphs: An Output Feedback Control Approach. *IFAC Proceedings Volumes* 47(3): 4681-4686 (2014).