Product Comparison Networks for Competitive Analysis of Online Word-of-Mouth

ZHU ZHANG, CHENHUI GUO, and PAULO GOES, The University of Arizona

Enabled by Web 2.0 technologies social media provide an unparalleled platform for consumers to share their product experiences and opinions—through word-of-mouth (WOM) or consumer reviews. It has become increasingly important to understand how WOM content and metrics thereof are related to consumer purchases and product sales. By integrating network analysis with text sentiment mining techniques, we propose product comparison networks as a novel construct, computed from consumer product reviews. To test the validity of these product ranking measures, we conduct an empirical study based on a digital camera dataset from Amazon.com. The results demonstrate significant linkage between network-based measures and product sales, which is not fully captured by existing review measures such as numerical ratings. The findings provide important insights into the business impact of social media and user-generated content, an emerging problem in business intelligence research. From a managerial perspective, our results suggest that WOM in social media also constitutes a competitive landscape for firms to understand and manipulate.

Categories and Subject Descriptors: I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Text analysis*; H.2.8. [**Database Management**]: Database Applications—*Data mining*

General Terms: Design, Performance

Additional Key Words and Phrases: Word-of-mouth, consumer reviews, text mining, network analysis, sentiment analysis, social media

ACM Reference Format:
Zhang, Z., Guo, C., and Goes, P. 2013. Product comparison networks for competitive analysis of online word-of-mouth. *ACM Trans. Manage. Inf. Syst.* 3, 4, Article 20 (January 2013), 22 pages.
DOI:http://dx.doi.org/10.1145/2407740.2407744

1. INTRODUCTION

Enabled by Web 2.0 technologies, social media provide an unparalleled platform for consumers to share their product experiences and opinions—through word-of-mouth (WOM) or consumer reviews. With this platform, WOM is generated in unprecedented volume and at great speed, and it creates unprecedented impacts on firm strategies and consumer purchase behavior [Dellarocas 2003; Godes et al. 2005; Zhu and Zhang 2010]. Chen and Xie [2008] argue that one major function of consumer reviews is to work as sales assistants to provide matching information, hence to help customers find products matching their needs. This suggests that considerable value of the consumer reviews, from a marketer's perspective, lies in the textual content.

However, when studying the sales impact of consumer reviews, many previous studies focus on the nontextual measures such as numerical ratings (e.g., Chevalier and Mayzlin [2006], Dellarocas et al. [2007], Duan et al. [2008], Zhu and Zhang [2010]). As

Authors' addresses: Z. Zhang, C. Guo, and P. Goes, Department of Management Information Systems, Eller College of Management, The University of Arizona, Tucson, AZ 85721; emails: zhuzhang@eller.arizona.edu, chguo@email.arizona.edu, pgoes@eller.arizona.edu.

DOI:http://dx.doi.org/10.1145/2407740.2407744

shown in the literature [Herr et al. 1991; Mizerski 1982], some types of WOM information tend to be more diagnostic than others. Such nuance and diagnostic information is quite possibly lost in translation when the consumer is asked to provide only a numerical rating, which is an overly succinct summary opinion. While the research gap is being quickly filled by recent studies on review content and how metrics thereof influence consumer purchases and product sales (e.g., Godes and Mayzlin [2004], Liu [2006], Ghose and Ipeirotis [2011], Zhang et al. [2012]), most prior research concentrates on single-entity (product)-oriented opinions; there has been no systematic effort in exploring competitive information existing in product reviews, in particular in the form of product comparisons. To address this important research issue, we propose product comparison networks as a novel construct that is computed from textual product reviews and captures the emergent structure of comparative opinions. We also design product ranking measures derived from such networks, which are not explored in previous literature. To test the validity of such metrics from a marketer's perspective, we conduct an empirical study based on the digital camera data from Amazon.com. Our results demonstrate a strong linkage between our proposed measures and product sales, in addition to the effect of consumer rating metrics and other text metrics in the literature.

The rest of the article is organized as follows. Section 2 reviews previous literature and presents the theoretical background of this research; Section 3 develops our proposed comparison networks for consumer reviews; Section 4 then presents the data, models, and findings of the empirical study; and Section 5 concludes with a summary and discussion of implications and future research opportunities.

2. LITERATURE REVIEW

In this section, we mainly review three bodies of related literature: (1) text sentiment mining, which is relevant to the computational basis of our research; (2) WOM as business intelligence for marketing and other management disciplines, which is the background for the business dimension; And (3) comparative opinions as well as their business implications, the focal area of our study.

2.1. Mining Sentiments in Text: Computational Efforts

Sentiment analysis has seen increasing attention by the computing community, mostly in the context of natural language processing [Liu 2012; Pang and Lee 2008]. Studies in this area typically focus on automatically assessing opinions, evaluations, speculations, and emotions in free text.

Sentiment analysis tasks include determining the level of subjectivity and polarity in textual expressions. Subjectivity assessment is to distinguish subjective text units from objective ones [Wilson et al. 2004]. Given a subjective text unit, we can further address its polarity [Nasukawa and Yi 2003; Nigam and Hurst 2004; Pang and Lee 2008; Pang et al. 2002; Turney 2001] and intensity [Pang and Lee 2005; Thelwall et al. 2010]. Affect analysis, which is similar to polarity analysis, attempts to assess emotions such as happiness, sadness, anger, horror, and so on [Mishne 2005; Subasic and Huettner 2001]. In recent years, sentiment analysis has been combined with topic identification for more targeted and finer-grained assessment of multiple facets in evaluative texts [Chung 2009; Hu and Liu 2004; Yang et al. 2010].

There are in general two approaches to text sentiment analysis. The first method is to initially employ lexicons and predefined rules to tag sentiment levels of words/phrases [Mishne 2006] in text and then aggregate to larger textual units [Li and Wu 2010; Liu et al. 2005]. Linguists have compiled several lexical resources for sentiment analysis, such as SentiWordNet [Esuli and Sebastiani 2006]. Lexicons can be enriched by linguistic knowledge [Subrahmanian and Reforgiato 2008; Zhang et al.

2009] or derived from text corpora statistically [Turney 2001; Yu and Hatzivassiloglou 2003].

As compared with the lexicon-based approach, a learning-based approach aims at finding patterns from precoded text snippets using machine learning techniques [Dave et al. 2003], including probabilistic models [Hu and Li 2011; Liu et al. 2007], Support Vector Machines [Airoldi et al. 2006; Pang et al. 2002], AdaBoost [Wilson et al. 2009], Markov Blankets [Airoldi et al. 2006], and so forth. To build effective models, various linguistic features, such as n-gram and POS tags [Zhang et al. 2009], and feature selection techniques [Abbasi et al. 2011] have been used to capture subtle and indirect sentiment expressions in context and to align with application requirements [Wiebe et al. 2004]. There have been effective sentiment analysis tools, such as Opinion Finder [Wilson et al. 2005], developed based on previous research. In a learning-based approach, training classifiers generally requires manually coded data aligned with target applications, at word/phrase [Wilson et al. 2009], sentence [Boiy and Moens 2009], or snippet [Hu and Li 2011; Liu et al. 2007] levels. Since manual coding of training data is typically labor-intensive, one can also first assess sentiments in smaller textual units with learning-based models and then aggregate to larger ones [Das and Chen 2007; Yu and Hatzivassiloglou 2003].

Sentiment analysis has been applied to summarizing people's opinions in news articles [Yi et al. 2003], political speeches [Thomas et al. 2006], and Web contents [Efron 2004]. Recent studies have extended sentiment and affect analysis to Web 2.0 contents, such as blogs and online forums [Li and Wu 2010; Liu et al. 2007]. In particular, due to their obviously opinionated nature, consumer reviews on products and services have received much attention from sentiment mining researchers [Pang et al. 2002; Turney 2001]. Several studies evaluated sentiment analysis methods on movie or product review corpora [Hu and Li 2011; Kennedy and Inkpen 2006]. There also have been system-driven efforts to include sentiment analysis in purchase decision making. For instance, Popescu and Etzioni [2005] introduce an unsupervised system to extract important product features for potential buyers. Liu et al. [2005] propose a framework that compares consumer opinions on competing products. These efforts can serve as preliminary computational infrastructure for social-media-driven business intelligence, which will be reviewed in the next section.

2.2. WOM as Business Intelligence: Textual Metrics

The emergence of social media promotes collective intelligence among online users [Surowiecki 2005], which affects individuals' decisions and influences organizations' operations. WOM in social media has started to gain much attention from business managers as an emerging type of business intelligence [Chen 2010].

In the finance literature, for example, social media have been used as indicators of public/investor perception of the stock market. Through manually extracting the "whisper forecasts" in forum discussions, Bagnoli et al. [1999] found that unofficial forecasts are actually more accurate than professional analysts' forecasts in predicting stock trends. Tumarkin and Whitelaw [2001] found that days of abnormally high message activity in stock forums coincide with abnormally high trading volumes, when online opinion is also correlated with abnormal returns. After applying the Naive Bayes algorithm to classify the sentiments of stock-related forum messages into three rating categories (bullish, bearish, or neither), Antweiler and Frank [2004] found that overall sentiments help predict market volatility and that variance of the rating categories is associated with increased trading volume. Das et al. [2005] and Das and Chen [2007] introduced more statistical and heuristic methods to assess forum message (and news) sentiments. They showed that both message volume and message sentiment are significantly correlated with stock price, but a coarse measure of disagreement is not significantly correlated with either market volatility or stock price.

A number of marketing studies have examined how WOM in social media influences product sales (e.g., Chen et al. [2011], Chevalier and Mayzlin [2006], Godes and Mayzlin [2004], Liu [2006], Zhang [2008], Zhu and Zhang [2010]); which is very important for marketing managers. Some of the most commonly studied WOM measures in this body of literature include volume (the amount of WOM information) and valence (overall sentiments of WOM, positive or negative). For instance Liu [2006] shows the volume of WOM has an explanatory power for movie box office revenue. Differently, Godes and Mayzlin [2004] find the dispersion of conversation volumes across communities, instead of the volume itself, has an explanatory power in TV viewership. Chevalier and Mayzlin [2006] demonstrate that the valence of consumer ratings has a significant impact on book sales at Amazon.com. Zhang [2008] demonstrates a positive correlation of usefulness-weighted review valence with product sales across four product categories. In addition, Forman et al. [2008] find that the disclosure of reviewer identity information and a shared geographical location between the reviewer and customer increases online product sales, therefore suggesting the potential impact of some interesting factors beyond the reviews themselves.

While none of these measures are constructed directly from the textual content of consumer reviews, some recent studies have made attempts to exploit textual information and have shown mixed results. Liu [2006] studied the sales effect of review sentiment valence based on human coding of the review text. Liu et al. [2010] examined review sentiment valence and subjectivity with text-mining techniques. In both studies, none of these content-based review measures is found to have an impact on or a correlation with product sales. Ghose and Ipeirotis [2011] found significant effects of review subjectivity on the sales of audio-video players but not digital cameras or DVDs. Most recently, Zhang et al. [2012] showed a significant impact of sentiment divergence on product sales.

2.3. Comparative Opinions and Competitive Intelligence: Research Gap

Almost all studies reviewed in the preceding, computing- or business-driven, focus on opinions oriented towards a single entity (e.g. a product). On the other hand, comparisons are often convincing methods of evaluation in human communication and social media, which lend themselves to further investigation.

The computational literature in this area is very limited. Jindal and Liu [2006a] use a combination of class sequential rules and Naïve Bayes classifiers to identify comparative sentences in text; they [Jindal and Liu 2006b] also use label sequential rules to further extract comparative relations. As purely computational studies, both fail to consider business implications.

On the competitive intelligence front, Xu et al. [2011] propose a conditional random field (CRF)-based method to extract comparative relations between products from online customer reviews. Their graphical model relies on the assumption that the product names in comparisons have to cooccur within the same sentence. One practical limitation of this method is that such fine-granularity cooccurrences are not often found in real online customer reviews. More importantly, the comparative relations are in isolation from each other, thus do not constitute a coherent picture; although the authors formulate the problem in the context of competitive intelligence, they do not establish any linkage between comparative opinions and any business variables. Netzer et al. [2012] also apply text mining methods to build graphs based on cooccurrence of product name entities. They construct an undirected network for analyzing relations among different products and exploring market structure of sedan cars and diabetes drugs. However, since the undirected topology does not capture comparative relations (which

Table I. A Brief Summary of Related Studies

Study	Network Perspective	Comparison Identification		BI Implication
		Method	Level	
Jindal and Liu 2006a	Candidate links	ML (CSR + NB)	sentence	N/A
Jindal and Liu 2006b	Directed links	ML (LSR)	sentence	N/A
Xu et al. 2011	Directed links	ML (CRF)	sentence	Little (vision only)
Netzer et al. 2012	Undirected links + nodes	Heuristic TM	message	Market structure
Current research	Directed links + nodes	Heuristic TM + ML	message	Market structure + product ranking
(TM = Text Mining; ML = Machine Learning)				

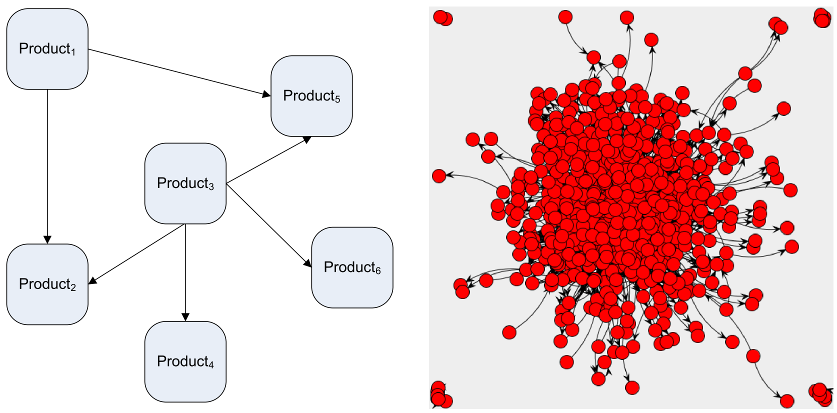

Fig. 1. A product comparison network (conceptual and actual).

are by nature directional), the authors emphasize structural-level findings based on correlation measures but fail to derive predictive power on individual nodes.

A brief summary of related studies (including our current research) is presented in Table I. We capitalize on both text sentiment mining and network analysis techniques to address significant gaps in previous research. More specifically, we efficiently use directional comparative opinions to compute product comparison networks that are capable of implying product sales.

3. RESEARCH DESIGN: PRODUCT COMPARISON NETWORKS

Our literature review has revealed that the community is increasingly paying attention to quantitative measures of consumer-generated product reviews and their influence on product sales. While there has been a considerable amount of research on WOM in social media and its implications for business intelligence and e-commerce applications, the extant early attempts in this area mainly focused on single-entity-oriented evaluative opinions in WOM (such as numerical ratings and polarity in text). Little attention has been paid to the understanding and modeling of competitive opinions, how products compare to each other according to consumers, as expressed in product reviews.

Inspired by both the success and limitations of previous research, we propose product comparison networks as a new computational construct, by exploiting comparative sentiments in social media and modeling them as a network-based competitive landscape. More specifically, each node in the network represents a product; a directed link indicates a comparative relationship between two products. Please see Figure 1 for illustration of product comparison networks. We also investigate how product ranking measures derived from such networks affect product sales, in addition to all existing review metrics in literature. As compared with previous measures, (1) our metrics are designed to capture comparative opinions (instead of single-entity-oriented valence) in product reviews, which adventure into an unexplored frontier in sentiment analysis; and (2) we focus on their impact on product sales, which has significant implications for WOM-driven marketing research.

The semantics of nodes in a product comparison network is relatively simple: each node corresponds to a product. We now discuss the link induction algorithm in detail. Generally, given an ordered product pair $<$ p1, p2 $>$, and the linguistic context, L $<$ p1, p2 $>$, in which they cooccur, one can derive the comparison relationship

$$C < p1, p2 >= f(L < p1, p2 >).$$

In this research, we define L<p1, p2>as a sentence in the review text for p1 along with a mention of p2; and f as a polarity classifier. Each sentence s is mapped into a comparison tuple t = {p1, p2, pScore, nScore}, where pScore is the positivity score of s and nScore the negativity score, determined by a polarity classifier. A positivity-dominated s implies p1's superiority to p2, and vice versa. The rough intuition behind our various network specifications is that if the linguistic context s conveys a considerable amount of dominating positivity, a directed link from p2 to p1 is produced (as if p2 casts a positive vote for p1); and vice versa.[1] This is similar to the hyperlinking situation on the World Wide Web, where a hyperlink from one page to another is considered a positive reference or vote. Over such a network, we can compute the prestige of a node (in our case, a product) by deploying the well-known node-ranking algorithms, PageRank [Page et al. 1998] and HITS [Kleinberg 1999]. Such network-induced measures will be used as predictors to infer product sales (please see Section 4).

Suppose there are n comparison tuples (i denoting the i-th tuple) sharing the same product pair <p1, p2>, depending on how these tuples are used to derive directed edges, we consider the following different graph settings.

Single-link graph. An edge e from node p2 to p1 is introduced when $\sum_{i=1}^{n} \{pScore_i - nScore_i\} > 0$. An edge e from node p1 to p2 is introduced when $\sum_{i=1}^{n} \{pScore_i - nScore_i\} < 0$. All n tuples are aggregated to produce a single link.

Dichotomic-link graph. A Type I edge e from node p2 to p1 is introduced when $\sum_{i=1}^{n} \left\{ \left(pScore_i - nScore_i\right) 1_{(pScore_i - nScore_i > 0)} \right\} > 0$; a Type II edge e from node p1 to p2 is introduced when $\sum_{i=1}^{n} \left\{ \left(pScore_i - nScore_i\right) 1_{(pScore_i - nScore_i < 0)} \right\} < 0$. In this case, an indicator variable splits all tuples into two groups. One group contains all tuples with positive values of $pScore_i - nScore_i$, while the other group contains all tuples with $pScore_i - nScore_i$. There will be 1 or 2 directed links between two product nodes p1 and p2 (Type I, Type II, or both). The dichotomic-link graph is affected by both positive and negative opinions, and also applies aggregation over multiple comparison tuples.

[1]Notice that the polarity of a local context s (only containing p2) is actually not equivalent to the direction of a comparison relationship C<p1, p2>. However:

— Such instantiation of L is practically more feasible than that used by Xu et al. [2011] and Jindal and Liu [2006b], given that E-commerce customers do not often directly compare two entities or products in one sentence.

— Such operationalization is also linguistically justifiable, given that p1 is clearly the focus in the current discourse [Walker et al. 1998], i.e., the review.

— Computationally, this simple unsupervised approach is not only much cheaper than the sophisticated supervised learning models in Xu et al. [2011] and Jindal and Liu [2006b], but has also proven highly successful in building meaningful comparison networks (please see Section 4 for empirical results).

Multi-link graph. An edge e from node p2 to p1 is introduced when $pScore_i - nScore_i > 0$; an edge e from node p1 to p2 is introduced when $pScore_i - nScore_i < 0$. No aggregation is performed. For each comparison tuple, a link is introduced; the direction of the link is determined by the difference between pScore and nScore. Equal weights are given to all opinions/links. Finally, there will be n directed links between two product nodes, p1 and p2.

In the following we also define the corresponding weighted versions of the graphs, where the links are weighted by the sentiment strength of the comparison relationship.

Single-link graph. An edge e from node p2 to p1 is introduced when $\sum_{i=1}^{n}\{pScore_i - nScore_i\} > 0$; an edge e from node p1 to p2 is introduced when $\sum_{i=1}^{n}\{pScore_i - nScore_i\} < 0$. All n tuples are aggregated to produce a single link with a weight.

$$\left|\sum_{i=1}^{n}\{pScore_i - nScore_i\}\right| \Big/ \sum_{i=1}^{n} 1_{(pScore_i - nScore_i \neq 0)} \,.$$

Dichotomic-link graph. A Type I edge e from node p2 to p1 is introduced when $\sum_{i=1}^{n}\left\{(pScore_i - nScore_i)1_{(pScore_i - nScore_i > 0)}\right\} > 0$; a Type II edge e from node p1 to p2 is introduced when $\sum_{i=1}^{n}\left\{(pScore_i - nScore_i)\, 1_{(pScore_i - nScore_i < 0)}\right\} < 0$. The weight of the Type I edge is determined by

$$\left|\sum_{i=1}^{n}\left\{(pScore_i - nScore_i)\, 1_{(pScore_i - nScore_i > 0)}\right\}\right| \Big/ \sum_{i=1}^{n} 1_{(pScore_i - nScore_i > 0)} \,.$$

While the weight of Type II edge is determined by

$$\left|\sum_{i=1}^{n}\left\{(pScore_i - nScore_i)\, 1_{(pScore_i - nScore_i < 0)}\right\}\right| \Big/ \sum_{i=1}^{n} 1_{(pScore_i - nScore_i < 0)} \,.$$

Multi-link graph. An edge e from node p2 to p1 is introduced when $pScore_i - nScore_i > 0$; an edge e from node p1 to p2 is introduced when $pScore_i - nScore_i < 0$. For each edge, the weight is determined by

$$|pScore_i - nScore_i|.$$

We further illustrate the different link settings in product comparison networks through the following example. Suppose we have two products p1="Canon PowerShot S70" and p2="Canon PowerShot S95." From text mining the reviews of these products we derive five comparison tuples t = {p1, p2, pScore, nScore} between p1 and p2, as follows.

t1={p1,p2, 5,1} t2={p1,p2,4,3} t3={p1,p2, 1,2} t4={p1,p2,1,3} t5={p1,p2,1,1},

where the positivity and negativity scores are obtained from a polarity classifier and reported in a 5-point scale (from 1 to 5). We use $pScore_i - nScore_i$ to represent the strength associated with comparison tuple i. For these 5 tuples we get 4, 1, –1, –2, and 0. We drop the 0-scored t5 and use the other four tuples to determine directed network links. Figure 2 illustrates the network construction and the calculation of the weights when p1 and p2 are the only two nodes.

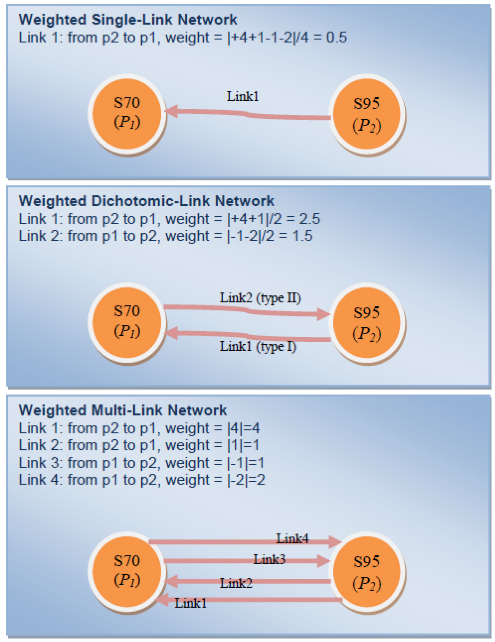

Fig. 2. Examples of link generation.

Table II. Data Summary Statistics

Data Summary	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9
# of Products	2,813	2,722	2,531	2,380	2,756	1,301	1,244	1,248	3,367
# of Reviews	64,994	66,395	63,014	62,655	67,200	50,826	52,074	47,975	76,731
# of Candidate Comp. Sentences	31,279	28,481	28,228	28,553	30,260	21,978	22,706	18,001	35,487
# of Good Comp. Sentences	17,317	15,261	14,665	14,721	16,256	10,533	11,091	7,556	19,833
# of Products in Comp. Network	805	742	720	705	781	472	471	397	911

4. EMPIRICAL STUDY

In this section, we use a longitudinal dataset collected from Amazon.com to empirically validate our proposed comparison network, in particular, the derived product ranking measures. We first describe our dataset, and then demonstrate the power of the comparison network with univariate correlation, regression (after controlling for exiting review measures in the literature), and prediction results.

4.1. Data Collection and Processing

We chose Amazon.com, a popular e-commerce site, as our data source, due to its offering of consumer ratings, reviews, and product sales information. A further technical advantage is the availability of an API, Amazon Web Services[2]. We managed to download 9 months' worth of data that contain from 1244 to 3367 digital cameras and correspondingly from 47,975 to 76,731 pieces of customer reviews and ratings. Table II

[2]http://aws.amazon.com/

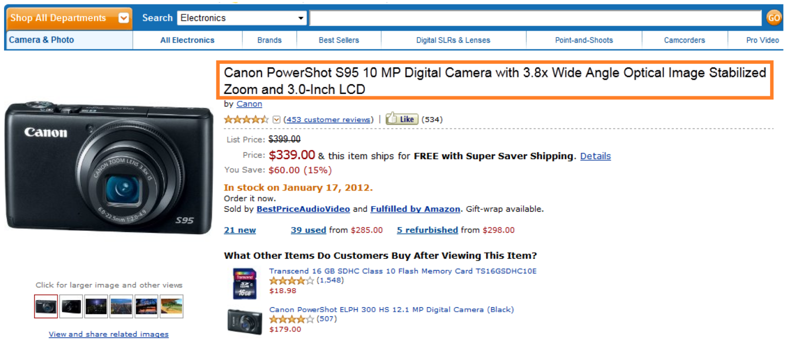

Fig. 3. Screenshot of Amazons retailing page.

provides a brief summary of our data. We process these reviews to build product comparison networks.

Textual Data Processing. The main task of textual data processing is to convert thousands of consumer-generated product reviews into comparison tuples t = {p1, p2, pScore, nScore}. Clearly, the first critical task is to identify comparison sentences. In order to have a comprehensive map of products, we start by building a catalog tree that contains all product-level entities. In Figure 3, we show a typical Amazon retailing page for a digital camera.

As highlighted in Figure 3, it is easy to find that the title of this product is "Canon PowerShot S95 10 MP Digital Camera with 3.8x Wide Angle Optical Image Stabilized Zoom and 3.0-Inch LCD," with Amazon's identification number (ASIN) "B003ZSHNGS." We identify product p1 by simply extracting the first three words of this title, namely "Canon" as a manufacturer level entity, "PowerShot" as a series level entity, and "S95" as a product level entity. Some product titles don't contain a series level entity, such as "Olympus TG-310 Tough 14 MP Digital Camera, 3.6× Wide Optical Zoom (28mm), 2.7" LCD, (Blue)." In this case, we only extract manufacturer level and product level entities ("TG-310"). Also, several simple heuristics are used here to deal with special cases and errors in product titles. By arranging three levels of entities as tree nodes, we finally construct a catalog hierarchy. We also assign an attribute to every node in the catalog tree, for which a "yes" denotes that the node is a product level entity and a "no" otherwise. Thus, any path from the root to a leaf node represents a particular product, such as "Canon PowerShot S95". It is worth mentioning that one product can have multiple ASINs associated to it, because of color options, special offers and so on. In such cases we aggregate all ASINs that refer to same products and only consider product comparisons at the aggregate product level. This Catalog Tree Construction algorithm is summarized in Figure 4 (minor noise reduction operations are not presented).

The tree-based catalog structure helps us store and manage product level entities (such as "S95") that are utilized in identifying p2 products in reviews. In order to get all comparison tuples with p1="Canon PowerShot S95," we search product reviews of "Canon PowerShot S95" for all product level entities except "S95." Figure 5 shows a review of "Canon PowerShot S95" that mentions another product called "Canon PowerShot S70." An entity matching algorithm (summarized in Figure 6) will go through this textual review and identify the latter (including possible variations such as "S70" or Canon S70"), which is a product level entity in the catalog tree. The algorithm also returns the sentence that contains "Canon PowerShot S70." Then the sentence is assigned a positivity score and a negativity score by SentiStrengh

```
Given a collection of product titles
Create catalog tree root node
  For each product ASIN i (i=1 to number of product ASINs), do
    Get the title of product I, and extract first, second, and third words
    If root doesn't have child node "first word,"
      Add node "first word" as a child node to root
    If node "first word" doesn't have child node "second word,"
      Add node "second word" as a child node to node "first word"
    If node "second word" doesn't have child node "third word,"
      Add node "third word" as a child node to node "second word"
    Create leaf node "product ASIN i"
    Add node "product ASIN i" as a child node to node "third word"
   End for
  For each non-leaf node in catalog tree, do
    Calculate SalesRank of aggregated product and set it as attribute "SalesRank"
    Set attribute "isprod" as "yes" if the word is a product level entity and "no" otherwise
  End for
  Return catalog tree
```

Fig. 4. The catalog tree construction algorithm.


891 of 908 people found the following review helpful:

Excellent image quality, full controls and pocketable

If you're looking for a pocketable camera that has reasonably high quality images, lets you control aperture, speed and focus and shoot in RAW format, this is it. I bought mine as an upgrade from a previous small but versatile camera, a Canon Powershot S70.

The Powershot S95 was introduced in August 2010 as a slight upgrade to the S90, which was widely...

Read the full review ›

Published 16 months ago by Michael Sandman

› See more 5 star, 4 star reviews

Fig. 5. An example of entity matching in a product review.

(Thelwall et.al. 2010),[3] an off-the-shelf sentiment strength detection tool. Finally we have a comparison tuple where p1="Canon PowerShot S95," p2="Canon PowerShot S70," a positivity score pScore, and a negativity score nScore. Again, we also use several simple heuristics in to deal with anomalies and typos in the text.

Network Construction Process. Given comparison tuples such as t = {p1, p2, pScore, nScore}, we build product comparison networks according to the previously defined network settings.[4] For benchmarking purposes, we also generate an undirected network, in which, no matter how many tuples there are between p1 and p2, we introduce an undirected link between the two nodes. This is similar to the cooccurrence network

[3]The algorithm assigns a positivity score (1 to 5) and negativity score (1 to 5) to a sentence, based on the following main elements and with a reported accuracy close to 97%. (1) A precoded sentiment word strength list; (2) machine-learned word strength adjustments; (3) a spelling correction component dealing with typos and informal writing; (4) a booster word list; (5) a negation word list; (6) an emoticon list; (7) tweaking related to punctuation marks.

[4]We exclude isolated nodes in the network. As a result, very new products that have not been compared in any consumer reviews are ignored in further analysis.

```
Given a collection of product titles, reviews and a catalog tree
Create empty set match with element structure (product A, product B, sentence);
  Traverse a catalog tree, and put all non-leaf nodes with attribute "isprod"="yes" into set "key entity"
  For each product ASIN i (i=1 to number of product ASINs), do
    Get the title of product ASIN i, and extract its first three words
    Find node "second word" in catalog tree, and
    If attribute "prod" of node "second word" is "yes", let "second word" be "Product A"
      Else find node "third word" in catalog tree, and
          If attribute "prod" of node "third word" is "yes", let "third word" be "Product A"
    For each review j of product ASIN i, do
      Tokenize review j in to words, and for each word w, do
        If set "key entity" contains word w, do
          Set "word w" as "Product B"
          Set the sentence in review j that contains word w as "sentence"
          Add a match element (product A, product B, sentence) into set match
        End if
      End for w
    End for j
   End for i
Return set match
```

Fig. 6. The entity matching algorithm.

introduced by Netzer et al. [2012]. Sentiment scores are not used at all, and the undirected link derived is unweighted (in other words, with a constant weight 1).

Measurement of Major Variables. SalesRank. We adopt SalesRank as a measure of sales performance of products, partially because it is the only sales performance measure available in Amazon's electronic market. On the other hand, previous literature also supports the legitimacy of using SalesRank as a proxy of demand [Archak et al. 2011; Chevalier and Goolsbee 2003; Chevalier and Mayzlin 2006; Oestreicher-Singer and Sundararajan 2012]. According to Chevalier and Goolsbee [2003], SalesRank follows a Pareto distribution (a power law). After log-transformation, the relationship between demand and SalesRank can be represented by a linear function.

$$\ln(\text{Demand}) = \alpha - \beta \ln(\text{SalesRank}),$$

where α is a constant, and β is the coefficient for $\ln(\text{SalesRank})$, which is empirically estimated to be between 0.9 and 1.3 in the book market [Chevalier and Goolsbee 2003]. Although earlier studies that apply this sales-demand transformation were all conducted in online book markets, Archak et al. [2011] used it for their study on the online digital camera and camcorder market.

Due to the ranking nature of SalesRank, it is important to keep in mind that smaller SalesRank values imply better sales performance. This also explains why, in our empirical studies, "if the correlation measure is negative, independent variable X has a positive influence on sales performance." It is worth mentioning that SalesRank data of corresponding products in Month t were collected in Month t+1. This time-lagging trick is in alignment with previous literature on marketing value of WOM [Duan et al. 2008; Liu 2006].

PageRank and HITS Authority. To quantify the effect of the proposed comparison networks, we compute prestige measures of products by using popular node scoring algorithms in directed networks. The PageRank score is computed by the PageRank algorithm [Page et al. 1998], implemented in the iGraph package in R. The Authority score (instead of the Hub score, due to the endorsement nature of incoming links in our comparison networks) is computed by the HITS algorithm [Kleinberg 1999] implemented in the JUNG toolkit [O'Madadhain et al. 2005].

Table III. Univariate Rank Correlation with SalesRank

Direction Setting	Network Setting	Ranking Measure	Kendalls τ	Spearmans ρ
Directed Graph	Single-link graph (720 nodes, 2905 links)	Unweighted HITS Authority	–.132**	–.194**
		Weighted HITS Authority	–.141**	–.206**
		Unweighted PageRank	–.169**	–.243**
		Weighted PageRank	–.171**	–.246**
	Dichotomic-link graph (720 nodes, 3626 links)	Unweighted HITS Authority	–.121**	–.179**
		Weighted HITS Authority	–.126**	–.184**
		Unweighted PageRank	–.175**	–.254**
		Weighted PageRank	–.184**	–.265**
	Multi-link graph (720 nodes, 6877 links)	Unweighted HITS Authority	–.122**	–.179**
		Weighted HITS Authority	–.136**	–.189**
		Unweighted PageRank	–.172**	–.254**
		Weighted PageRank	–.179**	–.260**
Undirected Graph	(Co-occurrence graph) (720 nodes, 3083 links)	Closeness Centrality	–.121**	–.179**
		Betweenness Centrality	–.128**	–.183**
	No Graph (720 products)	AvgRating (Star)	–.080**	–.109**
		NumReview (Volume)	–.222**	–.311**
		NumComparison	–.154**	–.221**

Note. *Correlation is significant at level of 0.05, **Correlation is significant at level of 0.01 (two-tailed).

Other Variables. As we known, product price plays an important role in shopping decisions. In online consumer markets, product sales performance may be more sensitive to price, because consumers gather price information more efficiently through online search. Therefore we include price in our models as a control variable. Besides, we also collect and derive other WOM measures, including average rating, number of reviews, number of comparisons, and centrality measures. Most of them have been used in previous literature (e.g., Chevalier and Mayzlin [2006], Archak et al. [2011]) in efforts of understanding the relationship between WOM and product sales, therefore they serve as valuable benchmark variables.

The average consumer rating (AvgRating) is an easy-to-compute, highly compact summary of overall consumer opinions, and a surrogate for isolated WOM evaluations. Another popular baseline variable in the literature is the volume of WOM, operationalized as number of reviews (NumReview). The number of times a product serves as a comparison target (NumComp) is a measure of a product's popularity in the context of cross-referencing. Centrality measures (closeness and betweenness) can also be computed (using the iGraph package again) on an undirected network. Similar to NumComp, these measures do not utilize directional sentiment information at all, but they illustrate the positions of products in market. Higher centrality measures may indicate importance or popularity of products. Considering NumComp and undirected centrality measures is helpful to justify the necessity of deriving directed comparison network structure from sentiment information.

4.2. Univariate Correlation Analysis and Results

To validate the power of the comparison-network-based scoring measures, we conduct a comprehensive rank correlation analysis between all product scoring measures (including those derived from directed graph, undirected graph and those computable without networks) and SalesRank (the dependent variable). For the sake of simplicity, we only present correlation results on Month 3's data, an average month. We compute the well-known Kendall's τ and Spearman's ρ rank correlations between the node scores (predicted value) and corresponding sales ranks (gold standard). The results are summarized in Table III.

As we can see from the correlation results, PageRank and HITS Authority derived from directed comparison networks consistently show strong negative correlation with

Table IV. Description of Variables and Descriptive Statistics (N = 4963)

Variable	Description	Min	Max	Mean	S. D.
$SalesRank_{it}$	Sales rank for product i at time t	1	443790	24506.02	29749.122
$AvgRating_{it}$	Average star rating of product i at time t	1	5	3.99	0.503
$NumReview_{it}$	Number of reviews on product i at time t	0	928	85.67	122.942
$NumComp_{it}$	Number of comparison to product i at time t	1	663	20.58	37.184
$Price_{it}$	Price of product i at time t	2	2200	299.82	168.844
$PageRank_{it}$	Weighted PageRank score of product i at time t	1.267×10^{-4}	2.119×10^{-2}	1.604×10^{-3}	2.245×10^{-3}
$Authority_{it}$	Weighted HITS Authority score of product i at time t	0	399.8	20.76	34.536
$Closeness_{it}$	Weighted Closeness centrality score of product i at time t	1.099×10^{-3}	0.455	0.116	0.0631
$Betweenness_{it}$	Weighted Betweenness centrality score of product i at time t	0	3.581×10^{4}	858.643	2128.844

product SalesRank, with Kendall's τ from –0.121 to –0.184 and Spearman's ρ from –0.179 to –0.265. No matter what network settings we apply, comparison network measures (PageRank or HITS Authority) always dominate our baseline variable AvgRating, which yields Kendall's τ at –0.080 and Spearman's ρ at –0.109. As a measure of WOM volume, NumReview also shows strong negative correlation with SalesRank. This result is in alignment with previous research [Archak et al. 2011; Chevalier and Mayzlin 2006; Liu 2006]. NumComp (Number of Comparisons), which is a measure reflecting products' cross-referencing popularity in customer reviews, with Kendall's τ at –0.154 and Spearman's ρ at –0.221, speaks more strongly than centrality measures, Closeness (Kendall's τ at –0.121 and Spearman's ρ at –0.179) and Betweenness (Kendall's τ at –0.128 and Spearman's ρ at –0.183). NumComp together with Betweenness and Closeness are by nature measures derived from an undirected graph; they underperform directed-graph-based measures because undirected cross-referencing (cooccurrence) does not reflect any directional opinions, i.e., customer preferences. Comparing different networking settings, we find that weighted graphs dominate unweighted ones. This implies that not only the direction, but also the magnitude of customer preferences in online Word-Of-Mouth, matters in influencing product sales. As to the different types of link settings, larger link magnitude (e.g., multilink) generally performs slightly better than smaller magnitude (e.g., single-link), due to the richness of information captured.

4.3. Regression Analysis and Results

To further validate the marginal power of the comparison-network-based ranking measures on top of existing product review metrics in the literature, we build a set of linear regression models on panel data. For the sake of simplicity, we only use the PageRank and Authority scores computed from the single-link, weighted version of the comparison network. Because of the wide dispersion of the values with several variables, we conduct log-transformation when appropriate. Before conducting the regression analysis, we explore the panel dataset and report results of the descriptive statistical analysis in Table IV and the correlations among variables of interest in Table V.

Then, we build linear models with control variables and dummy variables to estimate the power of the newly-built measure based on the 9-month data. In order

Table V. Correlation Triangle for Key Variables

	1.	2.	3.	4.	5.	6.	7.	8.
1.LogSalesRank								
2. AvgRating	–.069**							
3. LogPrice	.097**	.264**						
4. LogNumReview	–.171**	.233**	.089**					
5. LogNumComp	–.142**	.249**	.248**	.443**				
6. LogCloseness	–.153**	.082**	.050**	.167**	.148**			
7. LogBetweenness	–.120**	.197**	.170**	.458**	.834**	.128**		
8. LogAuthority	–.137**	.220**	.191**	.508**	.524**	.220**	.466**	
9. LogPageRank	–.270**	.190**	.133**	.489**	.689**	.229**	.650**	.636**

to mitigate the influence of unobservable variables, we introduce brand dummies and time dummies, similar to Duan et al. [2008] and Li and Hitt [2010]. The reason for including brand dummies is that we assume there exists unobservable heterogeneity across different brands of digital cameras, especially when marketers from different manufacturers promote products on Amazon's retailing channel differently. Time dummies help control some of the time variant heterogeneity that can be another potential source of estimation bias. Given our short panel dataset, with 9 time periods (which contains 983 digital cameras that belong to 27 different brands), it is not feasible to estimate models with product-level dummies efficiently since we lose many degrees of freedom. For comparison purposes, we build 8 least square dummy variable models, all of which have an intercept, a group of main regressors, 26 brand dummies, and 8 time dummies. We also conducted a Hausman Specification Test between the pooled OLS model and the LSDV model for each regression model [Hausman 1978]. All null hypotheses are rejected (with p-value < 0.001), indicating that LSDV models are preferred over simple pooled OLS models. These results strongly support the merit of models with brand and time dummies. The regression model specification is shown in the following.

Model 1.

$$\begin{aligned}\mathrm{LogSalesRank}_{i,t} = {} & \alpha + \beta_1 \mathrm{AvgRating}_{i,t-1} + \beta_2 \mathrm{LogNumReview}_{i,t-1} + \beta_3 \mathrm{LogPrice}_{i,t-1} \\ & + \lambda_1 \mathrm{BrandDummies}_i + \lambda_2 \mathrm{TimeDummies}_t + \varepsilon_{i,t}\end{aligned}$$

Model 2.

$$\begin{aligned}\mathrm{LogSalesRank}_{i,t} = {} & \alpha + \beta_1 \mathrm{AvgRating}_{i,t-1} + \beta_2 \mathrm{LogNumReview}_{i,t-1} + \beta_3 \mathrm{LogPrice}_{i,t-1} \\ & + \beta_4 \mathrm{LogNumComp}_{i,t-1} + \lambda_1 \mathrm{BrandDummies}_t + \lambda_2 \mathrm{TimeDummies}_1 + \varepsilon_{i,t}\end{aligned}$$

Model 3.

$$\begin{aligned}\mathrm{LogSalesRank}_{i,t} = {} & \alpha + \beta_1 \mathrm{AvgRating}_{i,t-1} + \beta_2 \mathrm{LogNumReview}_{i,t-1} + \beta_3 \mathrm{LogPrice}_{i,t-1} \\ & + \beta_4 \mathrm{LogNumComp}_{i,t-1} + \beta_5 \mathrm{LogCloseness}_{i,t-1} \\ & + \beta_6 \mathrm{LogBetweenness}_{i,t-1} + \lambda_1 \mathrm{BrandDummies}_i + \lambda_2 \mathrm{TimeDummies}_t \\ & + \varepsilon_{i,t}\end{aligned}$$

Model 4.

$$\begin{aligned}\mathrm{LogSalesRank}_{i,t} = {} & \alpha + \beta_1 \mathrm{AvgRating}_{i,t-1} + \beta_2 \mathrm{LogNumReview}_{i,t-1} + \beta_3 \mathrm{LogPrice}_{i,t-1} \\ & + \beta_4 \mathrm{LogNumComp}_{i,t-1} + \beta_5 \mathrm{LogCloseness}_{i,t-1} \\ & + \beta_6 \mathrm{LogBetweenness}_{i,t-1} \\ & + \beta_7 \mathrm{LogAuthority}_{i,t-1} + \lambda_1 \mathrm{BrandDummies}_i + \lambda_2 \mathrm{TimeDummies}_t + \varepsilon_{i,t}\end{aligned}$$

Model 5.

$$\begin{aligned}\mathrm{LogSalesRank}_{i,t} = {} & \alpha + \beta_1 \mathrm{AvgRating}_{i,t-1} + \beta_2 \mathrm{LogNumReview}_{i,t-1} + \beta_3 \mathrm{LogPrice}_{i,t-1} \\ & + \beta_4 \mathrm{LogNumComp}_{i,t-1} + \beta_5 \mathrm{LogCloseness}_{i,t-1} \\ & + \beta_6 \mathrm{LogBetweenness}_{i,t-1} \\ & + \beta_7 \mathrm{LogPageRank}_{i,t-1} + \lambda_1 \mathrm{BrandDummies}_i + \lambda_2 \mathrm{TimeDummies}_t + \varepsilon_{i,t}\end{aligned}$$

Table VI. Results of Linear Regression Models with Brand and Time Dummies

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Intercept	9.570** (.830)	9.439** (.832)	9.592** (.855)	9.504** (.856)	6.697** (.890)	6.510** (.893)	6.875** (.862)	6.621** (.865)
AvgRating	–.018 (.052)	–.016 (.052)	–.013 (.052)	–.013 (.052)	–.032 (.052)	–.034 (.052)	–.026 (.052)	–.031 (.052)
Log NumReview	–.150** (.020)	–.130** (.022)	–.125** (.022)	–.110** (.023)	–.068** (.023)	–.080** (.023)	–.048* (.022)	–.065** (.023)
Log Price	.468** (.045)	.488** (.046)	.484** (.046)	.490** (.046)	.465** (.045)	.457** (.045)	.496** (.044)	.465** (.045)
Log NumComp		–.050* (.024)	–.011 (.035)	–.001 (.036)	.075* (.036)	.068+ (.036)		.089** (.027)
Log Close			.042 (.070)	.068 (.071)	.042 (.069)	.010 (.070)		
Log Between			–.024 (.015)	–.022 (.015)	.008 (.015)	.009 (.015)		
Log Authority				–.018* (.009)		.023* (.010)		
Log PageRank					–.312** (.030)	–.344** (.033)	–.261** (.026)	–.308** (.029)
Adjusted R^2	.149	.150	.150	.151	.168	.169	.167	.168
N	4963	4963	4963	4963	4963	4963	4963	4963

Dependent variable is LogSalesRank in next time period.
Coefficients of brand and time dummies are not reported.
Significance Level: $^{+}p<0.1$, $^{++}p<0.05$, $^{*}p<0.01$, $^{**}p<0.001$.

Model 6.

$$\begin{aligned}\text{LogSalesRank}_{i,t} = {} & \alpha + \beta_1\text{AvgRating}_{i,t-1} + \beta_2\text{LogNumReview}_{i,t-1} + \beta_3\text{LogPrice}_{i,t-1} \\ & + \beta_4\text{LogNumComp}_{i,t-1} + \beta_5\text{LogCloseness}_{i,t-1} \\ & + \beta_6\text{LogBetweenness}_{i,t-1} \\ & + \beta_7\text{LogAuthority}_{i,t-1} + \beta_8\text{LogPageRank}_{i,t-1} + \lambda_1\text{BrandDummies}_i \\ & + \lambda_2\text{TimeDummies}_t + \varepsilon_{i,t}\end{aligned}$$

Model 7.

$$\begin{aligned}\text{LogSalesRank}_{i,t} = {} & \alpha + \beta_1\text{AvgRating}_{i,t-1} + \beta_2\text{LogNumReview}_{i,t-1} + \beta_3\text{LogPrice}_{i,t-1} \\ & + \beta_4\text{LogPageRank}_{i,t-1} + \lambda_1\text{BrandDummies}_i + \lambda_2\text{TimeDummies}_t + \varepsilon_{i,t}\end{aligned}$$

Model 8.

$$\begin{aligned}\text{LogSalesRank}_{i,t} = {} & \alpha + \beta_1\text{AvgRating}_{i,t-1} + \beta_2\text{LogNumReview}_{i,t-1} + \beta_3\text{LogPrice}_{i,t-1} \\ & + \beta_4\text{LogNumComp}_{i,t-1} + \beta_5\text{LogPageRank}_{i,t-1} + \lambda_1\text{BrandDummies}_i \\ & + \lambda_2\text{TimeDummies}_t + \varepsilon_{i,t}\end{aligned}$$

Coefficients of variables of interest are presented in Table VI, as well as standard errors of LSDV estimation and Adjusted R squared. As we can clearly see, Model 1 shows the coefficients of three benchmark variables, namely AvgRating, LogNumReview, and LogPrice, on LogSalesRank. This result is in alignment with previous studies and common sense. Products with higher WOM volume are more likely to achieve good sales performance, while higher price is harmful to sales. With the presence of brand dummies, coefficient of AvgRating on LogSalesRank is not significant, although correlation between the two is significant (see Table V). In Model 2, LogNumComp, the local popularity measure, has a moderate negative effect (p-value<0.01) on LogSalesRank, meaning that products attracting more comparisons in online WOM outperform those that seldom serve as comparison targets. Clearly, Model 1 (Adjusted R^2 = 0.149) and Model 2 (Adjusted R^2 = 0.150) are two baseline models without network-generated variables. Model 3 indicates that undirected network centrality measures, LogCloseness and LogBetweenness, have no power in explaining the LogSalesRank of products.

Models 4 through 8 show the merits of comparison-network-based measures. First, according to Model 4, LogAuthority has a moderate effect (Coefficient = –0.018, p-value < 0.01) on LogSalesRank, although R^2 doesn't change a lot when it enters the regression model (from 0.150 in Model 3 to 0.151 in Model 4). Second, compared with LogAuthority, LogPageRank (in Model 5, Coefficient = –0.312, p-value < 0.001) has much greater effect on LogSalesRank, and enhances R^2 from 0.151 to 0.168. The difference between these two directed-network-based measures is possibly due to HITS' susceptibility to link noise [Langville and Meyer 2006]. Then, in Model 6, the estimated coefficient of LogPageRank (–0.344, p-value < 0.001) is greater than that of any other independent variable. More importantly, Model 6 clearly indicates that the LogPageRank score dominates all measures derived from the undirected network, which signifies the critical advantage of directed comparison networks over undirected cooccurrence networks [Netzer et al. 2012]. Finally, Model 7 and Model 8 show how much improvement LogPageRank can bring to R^2 when entering the two baseline models (Models 1 and 2). As shown, R^2 increases from 0.149 in Model 1 to 0.167 in Model 7, or from 0.150 in Model 2 to 0.168 in Model 8. The last two linear models nicely achieve parsimony, and show the merits of the comparison-network-based PageRank measure for competitive analysis.

Overall, with potential unobservable heterogeneity mitigated by introducing brand and time dummies, the consistent significance of the LogPageRank score across multiple models demonstrates its explanatory power, even though it does not have the highest univariate correlation with SalesRank (seen in Table III). Moreover, despite its moderate correlation with several other independent variables[5] (seen in Table V), it always explains a unique portion of the variance.

4.4. Predictive Modeling and Results

To further illustrate the value of the comparison network and the derived product ranking measures, we conduct a set of prediction experiments. Based on the preceeding regression analysis, we compare the following models, by recruiting only the salient and observable variables (brand- and time-fixed effects are considered unobservable and unmeasurable in the prediction context). Inspired by Archak et al. [2011], we also incorporate a Trend variable to capture the outside publicity of digital camera brands, by using Google Trends to generate a "search volume" measure for each brand.[6] To illustrate the predictive power, we apply these models using current-month variables to predict sales performance two-months-ahead, and compare the performance of these models in terms of mean squared error (MSE). The month-by-month 5-fold cross validation results are presented in Table VII.

Model 1'.

$$\begin{aligned}\mathrm{LogSalesRank}_{i,t} = {}& \alpha_{t-2} + \beta_1 \mathrm{AvgRating}_{i,t-2} + \beta_2 \mathrm{LogNumReview}_{i,t-2} + \beta_3 \mathrm{LogPrice}_{i,t-2} \\ & + \beta_4 \mathrm{Trend}_{i,t-2} + \varepsilon_{i,t}\end{aligned}$$

Model 2'.

$$\begin{aligned}\mathrm{LogSalesRank}_{i,t} = {}& \alpha_{t-2} + \beta_1 \mathrm{AvgRating}_{i,t-2} + \beta_2 \mathrm{LogNumReview}_{i,t-2} + \beta_3 \mathrm{LogPrice}_{i,t-2} \\ & + \beta_4 \mathrm{Trend}_{i,t-2} + \beta_5 \mathrm{LogNumComp}_{i,t-2} + \varepsilon_{i,t}\end{aligned}$$

[5]It is not a surprise since LogNumReview, LogNumComp, LogAuthority, and LogPageRank all capture different aspects of "popularity".

[6]The way we normalize the Trend variable is treating the maximum search volume of "Canon PowerShot" (the most popular brand) as 1 and proportionating the other values.

Table VII. Results of Predictive Modeling in Mean Squared Error

Month (t)	N	Model 1	Model 2	Model 3
3	642	2.7171	2.6900	2.6572
4	541	2.4074	2.3826	2.3322
5	555	2.2137	2.1625	2.1136
6	394	2.4180	2.4029	2.3526
7	338	2.5324	2.6056	2.4988
8	290	3.5612	3.5135	3.4930
9	342	3.5602	3.4810	3.4213
10	336	2.4975	2.5064	2.4736

Model 3'.

$$\begin{aligned} \text{LogSalesRank}_{i,t} = {} & \alpha_{t-2} + \beta_1 \text{AvgRating}_{i,t-2} + \beta_2 \text{LogNumReview}_{i,t-2} + \beta_3 \text{LogPrice}_{i,t-2} \\ & + \beta_4 \text{Trend}_{i,t-2} + \beta_5 \text{LogNumComp}_{i,t-2} + \beta_6 \text{LogPageRank}_{i,t-2} + \varepsilon_{i,t} \end{aligned}$$

As we can see, when compared with the baseline model 1', the introduction of the PageRank measure (model 3') leads to a consistent decrease in MSE, while the naive popularity measure NumComparison (model 2') yields much less improvement to the prediction performance. The same pattern holds for all months. In a parsimonious fashion, such observation further justifies the value of the comparison network in sales prediction tasks. In reality, sales prediction or forecasting needs to incorporate a large number of nontrivial variables that capture detailed information of products and seasons and so on, which are not available in this research.

4.5. Practical Implications

Besides the academic value shown here, the comparison network as a construct also has important practical implications for firms.

First, as we have seen in Figure 1, it is a good visualization tool for the competitive landscape, typically in a coherent industry sector. Figure 7 shows the single-link directed network, computed from Month 3 data and reduced to only display the top-selling products. "Canon SD 780 IS" is the best-selling product and does get the highest PageRank score in the network; "Canon SD 1200 IS" ranks #2 in sales but only gets a moderate PageRank score. Despite the imperfect correlation, an emergent picture depicting complex interproduct comparisons in the digital camera sector is clearly fleshed out.

Second, it is also a good decision-support tool for business executives. For example, if Nikon management is not content with the sales performance of "Nikon S70", they may observe that Figure 7 suggests unfavorable consumer opinions relative to "Canon A1100 IS" and therefore hints, at the need for focused product design and improvement effort. Furthermore, if manipulating WOM is deemed legitimate and necessary, it makes sense for Nikon to exploit the comparison network and improve its product ranking. Specifically, effort could made to induce positive comparative opinions (visualized by the dashed link in Figure 7) of "Nikon S70" in favor of "Canon SD 780 IS" (the so-far most prestigious node in the network), in order to establish incoming links from the latter and hence acquire a higher PageRank score in the network.

4.6. Limitations

Like any other research endeavor, this study has limitations. We highlight several important ones in the following.

First, when constructing product-comparison networks, we face a technical challenge in deriving preference directions. Although our main contribution is not to design a new computational tool for link derivation per se, the algorithm used clearly has an

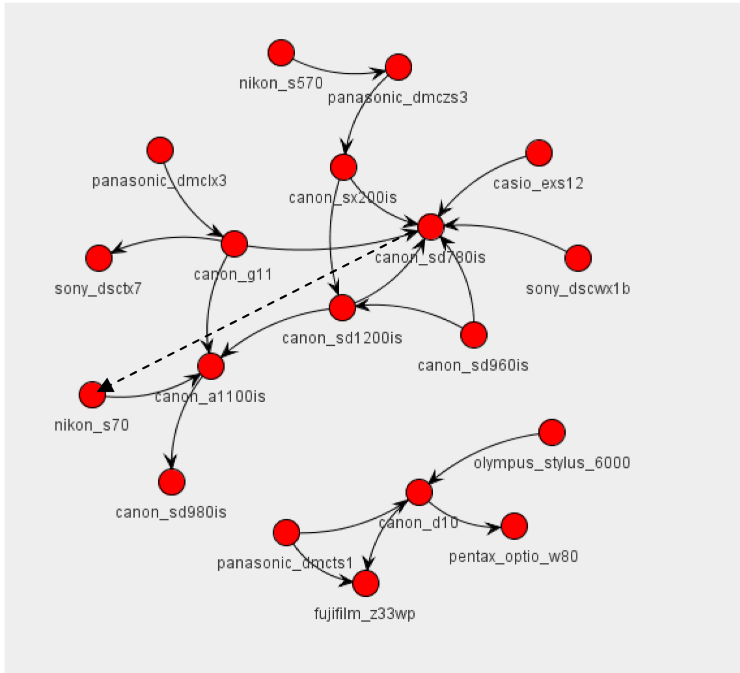

Fig. 7. A sample comparison network (reduced to only display the top-selling products).

impact on our results, since it determines the topology of the network, thus influencing network-based measures. In this study, we take a simplified approach to this problem, and we believe that the results of our study can be boosted by using more advanced text mining techniques. Second, since Amazon.com does not provide direct information about sales (e.g., market share), we use the logarithm of sales rank to approximate product sales. Although such treatment is in alignment with previous literature, we notice that there are measurement errors. Third, we do not focus on establishing causal relationships in the current study. Due to data limitations, we have little information about product-level characteristics and promotions. Therefore, the existence of missing/unobservable variables prevents us from specifying causality-based models. In typical causality-based demand models of differentiated products (e.g., discrete choice logit model), price is endogenous. Without proper instrument variables, we are not in a position to estimate the effect of price on demand. Potentially, future research can focus on modeling causality-based demand system of differentiated products with product comparison metrics.

5. CONCLUDING REMARKS

By integrating network analysis with text sentiment mining techniques, this study tackles an intriguing and challenging business intelligence problem. Specifically, upon proposing an innovative product comparison network, we found a significant effect of network-based measures on product sales performance, based on a large empirical dataset. Furthermore, this effect is not fully captured by traditional review measures.

These findings have important practical and theoretical implications. With the rapid expansion of social media, their influence on firm strategy and product sales has received a growing amount of attention from researchers and practitioners. Our results provide important insights for firms engaged in managing social media. The strong linkage between comparison-network-derived measures and product sales shows that firms can develop appropriate strategies to manage competitive information in review content. Our findings have not only recorroborated the value of textual consumer reviews, but also unveiled a totally new dimension in WOM-based business intelligence. Specifically, a product's sales performance not only depends on how it is perceived in isolation, but also how it is compared against its peers in a competitive landscape. This captures the notion of "competitive intelligence (CI)" [Prescott 1999], which, combined with competitive strategies [Porter 1980], ultimately leads to competitive advantage [Porter 1985] of a firm. While the Internet has been commonly agreed to be a key information source for CI [Teo and Choo 2001], social media sites are becoming an increasingly important platform (e.g., for monitoring public opinions and breaking news). Given the overwhelming amount of information available, it is critical to develop advanced analytic tools for competitive intelligence. We believe our research has clearly made contributions in this direction.

Our focus in the current article is to propose some valid network-based measures to capture the competitive landscape embodied in social media. Once we have demonstrated their validity in explanatory models, a probable application is to use such metrics (as well as other variables established in previous literature) to predict product sales. It is important to note that at the computational level, the currently used sentiment mining technique is very simple and parsimonious, in the sense that it only exploits shallow lexical information in the review text and transforms it into an approximated sentiment valance. Therefore room exists to further improve the link computation of product comparison networks, and consequently the network-based product ranking measures. Given the significant sales effects of these parsimonious and even naive metrics, we would expect a stronger effect from some more comprehensive metrics incorporating nuance information hidden in text (e.g., product-feature-level sentiments). For sales prediction purposes, future research can explore whether more sophisticated natural language processing and network modeling techniques would be helpful in better capturing this nuance information to make a better prediction.

Some recent research suggests the existence of self-selection biases in online product reviews (the early adopters' opinions may differ from those of the broader consumer population) and shows that consumers do not fully account for such biases when making purchase decisions [Li and Hitt 2008]. In light of such findings, opportunities exist for future research to strengthen our model by weighting or selecting product reviews (possibly based on recency or helpfulness).

When empirically testing the validity of our comparison networks, in alignment with previous literature we chose a setting in which both nontextual consumer ratings and textual content information were available. However, in reality, quite often only textual content is available for WOM in social media (e.g., message boards, chat rooms, and blogs). Future research can study how our textual metrics would explain and predict product sales or other financial performance in those settings. We suspect the impact of these measures might be even stronger there since nontextual, discrete rating information can lead to consumer herding behavior by ignoring detailed information from review content [Bikhchandani et al. 1992].

REFERENCES

Abbasi, A., France, S., Zhang, Z., and Chen, H. 2011. Selecting attributes for sentiment classification using feature relation networks. *IEEE Trans. Knowl. Data Eng. 23*, 3, 447–462.

Airoldi, E., Bai, X., and Padman, R. 2006. Markov blankets and meta-heuristics search: Sentiment extraction from unstructured texts. *Lect. Notes Artif. Intell. 3932*, 167–187.

Antweiler, W. and Frank, M. Z. 2004. Is all that talk just noise? The information content of Internet stock message boards. *J. Finance 59*, 3, 1259–1294.

Archak, N., Ghose, A., and Ipeirotis, P. G. 2011. Deriving the pricing power of product features by mining consumer reviews. *Manage. Sci. 57*, 8, 1185–1509.

Bagnoli, M., Beneish, M. D., and Watts, S. G. 1999. Whisper forecasts of quarterly earnings per share. *J. Account. Econ. 28*, 1, 27–50.

Bikhchandani, S., Hirshleifer, D., and Welch, I. 1992. A theory of fads, fashion, custom, and cultural-change as informational cascades. *J. Polit. Econ. 100*, 5, 992–1026.

Boiy, E. and Moens, M. F. 2009. A machine learning approach to sentiment analysis in multilingual web texts. *Inform. Retriev. 12*, 5, 526–558.

Chen, H. 2010. Business and market intelligence 2.0. *IEEE Intell. Syst. 25*, 6, 2–5.

Chen, Y. and Xie, J. 2008. Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Manage. Sci. 54*, 3, 477–491.

Chen, Y., Wang, Q., and Xie, J. 2011. Online social interactions: A natural experiment on word of mouth versus observational learning. *J. Marketing Res. 48*, 2, 238–254.

Chevalier, J. and Goolsbee, A. 2003. Measuring prices and price competition online: Amazon.com and BarnesandNoble.com. *Quant. Market. Econ. 1*, 2, 203–222.

Chevalier, J. A. and Mayzlin, A. 2006. The effect of word of mouth on sales: Online book reviews. *J. Market. Res. 43*, 3, 345–354.

Chung, W. 2009. Automatic summarization of customer reviews: An integrated approach. In *Proceedings of the 15th American Conference on Information Systems*.

Das, S., Martinez-Jerez, A., and Tufano, P. 2005. eInformation: A clinical study of investor discussion and sentiment. *Financ. Manage. 34*, 3, 103–137.

Das, S. R. and Chen, M. Y. 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Manage. Sci. 53*, 9, 1375–1388.

Dave, K., Lawrence, S., and Pennock, D. M. 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th International World Wide Web Conference*.

Dellarocas, C. 2003. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Manage. Sci. 49*, 10, 1407–1424.

Dellarocas, C., Zhang, X. Q., and Awad, N. F. 2007. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *J. Interactive Market. 21*, 4, 23–45.

Duan, W., Gu, B., and Whinston, A. B. 2008. Do online reviews matter? An empirical investigation of panel data. *Dec. Supp. Syst. 45*, 4, 1007–1016.

Efron, M. 2004. Cultural orientation: Classifying subjective documents by vocation analysis. In *Proceedings of the AAAI Fall Symposium on Style and Meaning in Language, Art, Music, and Design*.

Esuli, A. and Sebastiani, F. 2006. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation*, 417–422.

Forman, C., Ghose, A., and Wiesenfeld, B. 2008. Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Inform. Syst. Res. 19*, 3, 291–313.

Ghose, A. and Ipeirotis, P. G. 2011. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Trans. Knowl. Data Eng. 23*, 10, 1498–1512.

Godes, D. and Mayzlin, D. 2004. Using online conversations to study word-of-mouth communication. *Market. Sci. 23,* 4, 545–560.

Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., Pfeiffer, B., Libai, B., Sen, S., Shi, M., and Verlegh, P. 2005. The firm's management of social interactions. *Market. Lett. 16*, 3–4, 415–428.

Hausman, J. A. 1978. Specification tests in econometrics. *Econometrica 46*, 6, 1251–1271.

Herr, P. M., Kardes, F. R., and Kim, J. 1991. Effects of word-of-mouth and product-attribute information on persuasion: An accessibility-diagnosticity perspective. *J. Consumer Res. 17*, 4, 454–462.

Hu, M. and Liu, B. 2004. Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.

Hu, Y. and Li, W. 2011. Document sentiment classification by exploring description model of topical terms. *Comput. Speech Lang. 25*, 2, 386–403.

Jindal, N. and Liu, B. 2006a. Identifying comparative sentences in text documents. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development on Information Retrieval*.

Jindal, N. and Liu, B. 2006b. Mining comparative sentences and relations. In *Proceedings of the 21st National Conference on Artificial Intelligence*.

Kennedy, A. and Inkpen, D. 2006. Sentiment classification of movie reviews using contextual valence shifters. *Comput. Intell. 22*, 2, 110–125.

Kleinberg, J. 1999. Authoritative sources in a hyperlinked environment. *J. ACM. 46*, 5, 604–632.

Langville, A. N. and Meyer, C. D. 2006. *Google's PageRank and Beyond: The Science of Search Engine Rankings.* Princeton University Press, Princeton, NJ.

Li, N. and Wu, D. D. 2010. Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decis. Supp. Syst. 48*, 2, 354–368.

Li, X. and Hitt, L. M. 2008. Self-selection and information role of online product reviews. *Inform. Syst. Res. 19*, 4, 456–474.

Li, X. and Hitt, L. M. 2010. Price effects in online product reviews: An analytical model and empirical analysis, *MIS Quart. 34*, 4, 809–831.

Liu, B. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

Liu, B., Hu, M., and Cheng, J. 2005. Opinion observer: Analyzing and comparing opinions on the Web. In *Proceedings of the 14th International World Wide Web Conference*, 342–351.

Liu, Y. 2006. Word of mouth for movies: Its dynamics and impact on box office revenue. *J. Market. 70*, 3, 74–89.

Liu, Y., Chen, Y., Lusch, R., Chen, H., Zimbra, D., and Zeng, S. 2010. User-generated content on social media: Predicting new product market success from online word-of-mouth. *IEEE Intell. Syst. 25*, 6, 8–12.

Liu, Y., Huang, X., An, A., and Yu, X. 2007. ARSA: A Sentiment-aware model for predicting sales performance using blogs. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 607–614.

Mason, C. H. and Perreault, W. D. 1991. Collinearity, power, and interpretation of multiple-regression analysis. *J. Market. Res. 28*, 3, 268–280.

Mishne, G. 2005. Experiments with mood classification in blog posts. In *Proceedings of the 1st Workshop on Stylistic Analysis of Text for Information Access*.

Mishne, G. 2006. Multiple ranking strategies for opinion retrieval in blogs. In *Proceedings of the 15th Text Retrieval Conference*.

Mizerski, R.W. 1982. An attribution explanation of the disproportionate influence of unfavorable information. *J. Consumer Res. 9*, 3, 301–310.

Nasukawa, T. and Yi, J. 2003. Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd International Conference on Knowledge Capture*. 70–77.

Netzer, O., Feldman, R., Goldenberg, J., and Fresko, M. 2012. Mine your own business: Market structure surveillance through text mining. *Market. Sci. 31*, 3, 521–543.

Nigam, K. and Hurst, M. 2004. Towards a robust metric of opinion. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text*. 598–603.

Oestreicher-Singer, G. and Sundararajan, A. 2012. Recommendation networks and the long tail of economic networks. *MIS Quart. 36*, 1, 65–83.

O'Madadhain, J., Fisher, D., Smyth, P., White, S., and Boey, Y.-B. 2005. Analysis and visualization of network data using JUNG. *J. Statist. Softw.* VV 1–35.

Page, L., Brin, S., Motwani, R., and Winograd, T. 1998. The pagerank citation ranking: Bringing order to the web, Stanford InfoLab. http://ilpubs.stanford.edu:8090/422/.

Pang, B. and Lee, L. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. 115–124.

Pang, B. and Lee, L. 2008. Opinion mining and sentiment analysis. *Found. Trends Inform. Retrieval 2*, 1–2, 1–135.

Pang, B., Lee, L., and Vaithyanathan, S. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*. 79–86.

Popescu, A. M. and Etzioni, O. 2005. Extracting product features and opinions from reviews. In *Proceedings of Conference on Human Language Technology and Empirical Methods in Natural Language Processing*. 339–346.

Porter, M. E. 1980. *Competitive Strategy: Techniques for Analyzing Industries and Competitors.* Free Press, NY.

Porter, M. E. 1985. *The Competitive Advantage: Creating and Sustaining Superior Performance.* Free Press, NY.

Prescott, J. 1999. The Evolution of Competitive Intelligence, Designing a Process for Action. *APMP Professional J.* Spring.

Subasic, P. and Huettner, A. 2001. Affect analysis of text using fuzzy semantic typing. *IEEE Trans. Fuzzy Syst. 9*, 4, 483–496.

Subrahmanian, V. S. and Reforgiato, D. 2008. AVA: Adjective-verb-adverb combinations for sentiment analysis. *IEEE Intell. Syst. 23*, 4, 43–50.

Surowiecki, J. 2005. *The Wisdom of Crowds*. Anchor Books, New York.

Teo, T. S. H. and Choo, W. Y. 2001. Assessing the impact of using the Internet for competitive intelligence. *Inform. Manage. 39*, 1, 67–83.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. 2010. Sentiment strength detection in short informal text. *J. Amer. Soc. Inform. Sci. Tech. 61*, 12, 2544–2558.

Thomas, M., Pang, B., and Lee, L. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the Conference on Empirical Methods on Natural Language Processing*. 327–335.

Tumarkin, R. and Whitelaw, R. F. 2001. News or noise? Internet postings and stock prices. *Financ. Anal. J. 57*, 3, 41–51.

Turney, P. D. 2001. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. 417–424.

Walker, M., Joshi, A. K., and Prince, E. F. 1998. *Centering Theory in Discourse*. Clarendon Press, Oxford, UK.

Wiebe, J., Wilson, T., Bruce, T., Bell, M., and Martin, M. 2004. Learning subjective language. *Comput. Ling. 30*, 3, 277–308.

Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., and Patwardhan, S. 2005. OpinionFinder: A system for subjectivity analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*.

Wilson, T., Wiebe, J., and Hoffmann, P. 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Comput. Ling. 35*, 3, 399–433.

Wilson, T., Wiebe, J., and Hwa, R. 2004. Just how mad are you? Finding strong and weak opinion clauses. In *Proceedings of the 19th National Conference on Artificial Intelligence*. 761–769.

Yang, C., Tang, X., Wong, Y. C., and Wei, C.-P. 2010. Understanding online consumer review opinion with sentiment analysis using machine learning. *Pacific Asia J. Assoc. Inform. Syst. 2*, 3, 7.

Yi, J., Nasukawa, T., Bunescu, R., Niblack, W., and Center, C. 2003. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Proceedings of the 3rd IEEE International Conference on Data Mining*. 427–434.

Xu, K., Liao, S. S., Li, J., and Song, Y. 2011. Mining comparison opinions from customer reviews for competitive Intelligence. *Dec. Supp. Syst. 50*, 4, 743–754.

Yu, H. and Hatzivassiloglou, V. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 129–136.

Zhang, C., Zeng, D., Li J., Wang, F.-Y., and Zuo, W. 2009. Sentiment analysis of Chinese documents: From sentence to document level. *J. Amer. Soc. Inform. Sci. Tech. 60*, 12, 2474–2487.

Zhang, Z. 2008. Weighing stars: Aggregating online product reviews for intelligent e-commerce applications. *IEEE Intell. Syst. 23*, 5, 42–49.

Zhang, Z., Li, X., and Chen, Y. 2012. Deciphering word-of-mouth in social Media: Text-based metrics of consumer reviews. *ACM Trans. Manage. Inf. Syst. 3*, 1.

Zhu, F. and Zhang Z. 2010. Impacts of online consumer reviews on sales: The moderating role of product and consumer characteristics. *J. Market. 74*, 2, 133–148.

Received April 2012; revised August 2012, October 2012; accepted October 2012