Electronic Commerce Research
https://doi.org/10.1007/s10660-022-09534-y

A novel approach for product competitive analysis based on online reviews

Zhen He[1] · **Lu Zheng**[1] · **Shuguang He**[1]

Accepted: 18 January 2022

Abstract

Recently, online reviews have become a prevalent information source for competitive analysis because they provide rich information on the voices of customers. Based on online reviews, we propose a novel method named Integrated-Degree based K-shell decomposition (ID-KS) to conduct competitive analysis via product comparison networks. Under the consideration of feature differences among products, we apply text-mining approaches and ID-KS to convert online reviews into competitive insights including competitor identification, product comparison, product ranking, brand comparison and market-structure analysis. To validate the feasibility and the effectiveness of ID-KS, we demonstrate our approach in two cases, SUV cars and laptops, and compare it with state-of-the-art methods. The results show that ID-KS analyzes product comparison networks more effectively and properly, and it derives comprehensive comparative insights that are not fully captured by existing studies.

Keywords Competitive analysis · Text analysis · Latent dirichlet allocation · Sentiment analysis · K-shell decomposition

1 Introduction

Competitive analysis is vital for manufacturers competing in a fierce market [1]. It is helpful to reveal the strengths and weaknesses of products, and to help manufacturers improve product quality [2, 3]. However, traditional ways to obtain competitive information are usually through surveys and interviews which are time-lagging and insufficient [4, 5]. Recently, the surge of information technology provokes the explosive growth of social media data. As a typical social media data, online reviews contain sufficient and trustworthy information on products [5, 6]. Hence, online reviews have been widely used and are increasingly prevalent in studies for competitive

✉ Lu Zheng
luzheng@tju.edu.cn

1 College of Management and Economics, Tianjin University, Tianjin 300072, China

Published online: 14 February 2022

analysis given their availability. Extant studies mainly leverage product comparison networks to discover competitive relationships and to gather meaningful managerial insights for competitive analysis [1, 7–10]. These studies have proven the effectiveness of network analysis in product competitive analysis. But several difficulties still need to be overcomed when conducting competitive analysis via networks. First, almost all of the research focuses on comparative sentences to measure the co-occurrence of two products (or brands), which is used as the indicator of comparative relations. But the limited number of comparative sentences affects the accuracy and comprehensiveness of comparative results. According to the finding published by Yang et al., the percentage of reviews containing comparative sentences in the datasets that most studies used is fewer than 10%. Insufficient reviews with comparative sentences may increase the risk of biased reviews and cause misleading comparative results [9]. Therefore, how to overcome the restriction of rare reviews with comparative sentences becomes the first challenge when making a thorough comparative analysis. Second, to gain competitive advantages, more manufacturers choose to differentiate their product features to meet the diversified needs of customers. How to conduct competitive analysis with the consideration of product feature differences still needs further research. Third, popular approaches of Social Network Analysis (SNA) are used in extant research to analyze comparison networks. However, existing SNA methods are developed to solve the problems of indirect networks. They base on the degree, which means the number of links that a node has with other nodes. When applying these methods in direct networks, most extant studies just calculate the degree using the sum of difference of in-degree or out-degree, and then transfer direct networks into indirect networks [11]. Such overlooking of the direction of relations will cause the loss of much information when analyzing social networks. Moreover, when most nodes have the same in-degree and out-degree, the extant approaches will assume that these nodes have the same degree and then fail to distinguish them [12]. Therefore, an effective network analysis approach is needed for comparison network analysis.

To tackle these research gaps, we propose a novel approach named Integrated Degree based K-shell Decomposition (ID-KS) to make a competitive analysis. First, we assess the feature performance of products using Latent Dirichlet Allocation (LDA) and sentiment analysis. In this process, we consider feature differences and divide product features into common features and unique features. Common features are the features mentioned by most products' reviews, while unique features are only discussed in partial products' reviews. Then we obtain the total performance of a product by aggregating its feature performance and constructing comparative networks. To analyze comparative networks, we utilize ID-KS to derive thorough information on product comparison, competitor identification, product ranking, brand comparison and market structure analysis which are not addressed in previous works.

To summarize, the contributions of this work are three-fold. First, instead of mining comparative sentences, we use the whole online reviews to overcome the drawbacks of limited reviews with comparative sentences. Insights derived from the whole online reviews are unbiased and more thorough. Second, considering feature differences, we evaluate the product performance by assessing common feature performance and unique feature performance separately. To the best of our knowledge, this is the first attempt to consider feature differences when doing such competitive analysis. Third, we propose a novel method called ID-KS to analyze comparison networks. ID-KS is not only validated to outperform in experimental comparisons, but also provides more managerial implications than extant research.

The remainder of this paper is organized as follows. In Sect. 2, we conduct a comprehensive literature review of related works. In Sect. 3, we lay out the details of our method. Case studies are implemented to validate ID-KS's effectiveness in Sect. 4. Section 5 concludes our study and provides an overview of its limitations and opportunities for future work.

2 Literature review

2.1 Online reviews

As a prevalent way to gather product information, online reviews attract attention from both manufacturers and researchers. For the advantages of promptness and availability, online reviews have been utilized in various aspects including product design [13–17], product defect discovery [18–21], service quality measurement [22–24] and box-office prediction [25, 26].

Recently, researchers have attempted to exploit online reviews for competitive analysis and published many significant research findings. Some researchers focus on product ranking and utilize customer sentiment of reviews as the ranking criteria. Li et al. ranked products based on the product affinity, which is evaluated through online reviews [27]. Yang et al. integrated heterogeneous information from reviews and constructed a multi-product ranking system [9]. Liu et al. combined sentiment analysis with intuitionistic fuzzy set theory to rank products via online reviews [3]. Instead of product ranking, some researchers study product comparison. Wang and Wang compared two products and discovered product weaknesses [28]. Wang et al. analyzed two competitive products by finding differences in product features [29]. Given that comparing only two products will lead to the loss of much information, Nasr et al. proposed the product comparison matrix for multiple product comparisons [2]. Before comparing products, identifying which products are comparative is necessary. Hence, competitor identification becomes a research hotspot in competitive analysis. Jin et al. collected online reviews and identified comparative opinionated sentences for competitor identification [30]. Gao et al. utilized LDA to mine comparative relations and discover competitors in the restaurant industry [1]. All these studies above prove the effectiveness of online reviews in various research areas.

2.2 Social network analysis

The comparative network is a tool to present comparative relations and has been applied in competitive analysis by many studies. In comparative networks, nodes denote products while edges denote comparative relations. To analyze comparative networks, previous studies utilized lots of SNA methods. Degree is the primary one to discover essential nodes [31]. However, this method measures influence from the nodes' neighbors without considering nonadjacent nodes. Centrality is a series of methods to estimate node influence. It includes degree centrality, closeness centrality, betweenness centrality and so forth [31]. Except for degree and centrality, k-shell decomposition has been proved to be effective and accurate [12, 32]. K-shell decomposition is to partition networks based on the node location in the network. After decomposition, every node in the network is assigned a KS value. The larger the KS value, the more important the node. Classical k-shell decomposition performs well but it only deals with unweighted and undirected graphs [32]. Then Garas et al. developed weighted K-shell decomposition, which can be used in weighted and undirected graphs [33]. However, neither of these two approaches can discriminate nodes well when most nodes have the same degree. To overcome this shortcoming, Wang et al. [12] developed an improved K-shell decomposition named KS-IF. Wei et al. [34] also improved the classical K-shell method and partition weighted networks based on potential edge weights. Despite the compelling performance of these two methods, they still cannot handle the weighted direct networks such as comparative relations networks and estimate influence accurately.

Researchers have increasingly leveraged the aforementioned advantages of comparative networks in competitive analysis because they describe complex comparative relations among multiple products clearly [9]. Network analysis enables researchers to gather more competitive insights from product comparison networks. Li et al. attempted to compare products using directed networks based on comparative sentences [7]. When proposing a network analysis framework for market-structure surveillance, Netzer et al. argued that "co-occurrence" is the indicator of comparative relations, and then used undirected networks to achieve the objectives of market structure analysis, brand comparison and product comparison [35]. Zhang et al. made great contributive researches on comparative networks. They developed three types of comparative networks to compare products under different circumstances [8]. Their methods then became an essential basis of the latest studies [1, 36]. All the studies mentioned above discover comparative relations from comparative sentences. However, Yang et al. argued that using comparative sentences hidden in online reviews may lead to biased results because of the negligible amount. To overcome the shortcomings of comparative sentences, they merged comparative sentences with comparative votes from third-party websites to establish comparative relations for multiple products [9]. Network-based competitive analysis has shown its outperformance. But it should be improved to address the following problems.

(1) The shortage of comparative sentences buried in online reviews may cause misleading results. Besides, the negligible number of reviews with comparative sentences cannot provide sufficient comparative relations for a comprehensive competitive analysis [9].
(2) When constructing comparative networks, few studies consider the competitive influence propagation which is important in network analysis. Most network-based methods only consider the direct comparative relations and omit the indirect relations which are created due to the competitive influence propagation.

Table 1 Summary of literature on the competitive analysis

Researches	Methods	Network types	Influence propagation	Feature difference	Data types	Managerial implication
Zhang et al. [16]	TA	–			OR	PC
Liu et al. [3]	IFST	–			OR	PR
Li et al. [17]	AffRank	–			Numeric data	PR
Ding et al. [12]	CF	–			Comments	PR
Valkanas et al. [18]	CMINER	–			OR	CI
Liu et al. [13]	TA + ML	–			CS	CI + PC
Netzer et al. [14]	Network	UN			CS	MSA + BC + PC
Li et al. [7]	Network	DN	√		CS	PC
Zhang et al. [8]	Network	UN + DN			CS	PR
Chen et al. [15]	Network	UN + DN	√		CS	PR
Yang et al. [9]	Network	UN + DN			CS + CV	PR
Gao et al. [1]	Network	UN + DN			CS	MSA + CI + PC
This study	Network	DN	√	√	OR	MSA + CI + BC + PC + PR

TA: Text Analytic; *ML*: Machine Learning; *IFST*: Intuitionistic Fuzzy Set Theory; *CF*: Collaborative filtering; *CS*: comparative sentences; *UN*: undirected network; *DN*: directed network; *OR*: Online reviews; *MSA*: Market Structure Analysis; *PR*: Product Ranking; *PC*: Product Comparison; *CV*: Comparative votes; *CI*: competitor identification; *BC*: brand comparison

Moreover, extant studies do not take product feature differences into account when comparing products. It is unreliable to compare products by assessing the performance of the same product features. Table 1 summarizes the research on competitive analysis via online reviews. From Table 1 we identify the research gaps and improve these gaps by (1) utilizing the whole reviews instead of reviews with comparative sentences in the competitive analysis, (2) involving competitive influence propagation in the construction of comparative networks, and (3) considering product feature differences in the competitive analysis. In contrast to existing studies, our work provides more managerial implications and helps manufacturers to get thorough insights about their products or brands.

2.3 LDA and sentiment analysis

Although online reviews have immense value, explosive growth in volume poses daunting challenges for manufacturers to process. It is necessary to help decision-makers digest such a colossal quantity of data and transfer them into knowledge without manual reading and tagging [37]. This demanding problem prompts the burgeoning development of text analysis techniques. One technique with significant performance is LDA. LDA transforms text data into several topics and topic-related words [38]. Since its convenience and reliability, LDA has been applied in various researches. Guo et al. collected reviews from travel websites and obtained related topics using LDA. By observing these topics, they discovered the aspects that customers care about [39]. Xiang et al. leveraged LDA to contrast the leading travel websites. Their research shows that bias exists on these websites [40]. Wang et al. fused structural data with unstructured data to detect insurance fraud. They combined LDA with deep learning and enhanced the performance of insurance fraud identification [41]. All of the aforementioned research verify that LDA is an effective topic model to process text data.

In addition to the opinions of reviews, review sentiment also attracts much attention from academic researchers. Because the sentiment of online reviews indicates the attitudes of reviewers, many studies use sentiment analysis to evaluate customer satisfaction and explore the underlying influencing factors of satisfaction [23, 24, 42–44]. Beyond satisfaction, sentiment analysis is extensively used in the research areas of decision-making [45], social network analysis [46] and product development [47]. Given that customer attitudes are an indicator of product performance, we perform sentiment analysis on online reviews to measure product performance.

3 Research methodology

3.1 Data preprocessing and feature extraction

The first step of our method is to preprocess data and extract product features used for comparison. Data preprocessing includes stop-word removal and special symbols removal. Given that a sentence usually contains a discussed object and a kind of sentiment [48], we conduct the sentence segment for each review. Then we extract product features for comparison. We assume the set of products is $\mathbf{P}$ and $\mathbf{P}=\{P_1, P_2,\ldots, P_m\}$, where m is the number of products. Considering that feature differences exist among products, we choose to derive common features and unique features separately.

LDA is a widely used topic model to extract topics from a large corpus [38]. We use LDA to extract topics and topic-related sentences for each product from its online reviews. We define the set of features extracted from P_i's reviews as $\mathbf{F}_i$. For each product, we extract its common features and unique features respectively. Common features are the features discussed by reviews of at least H products. H is the restriction parameter, and $0<H\leq m$. For example, when $H=15$, common features are features that are referred to in reviews of at least 15 products. Unique features are the features mentioned in the reviews of less than H products. We define the set of common features as $\mathbf{CF}$. As for P_i, its unique features are calculated by the difference set of $\mathbf{F}_i$ and $\mathbf{CF}$. We define the set of unique features for P_i as $\mathbf{UF}_i$. LDA provides not only the main topics of a large corpus but also topic-related sentences. Next, we evaluate product performance by measuring the sentiment of these topic-related sentences.

3.2 Performance evaluation

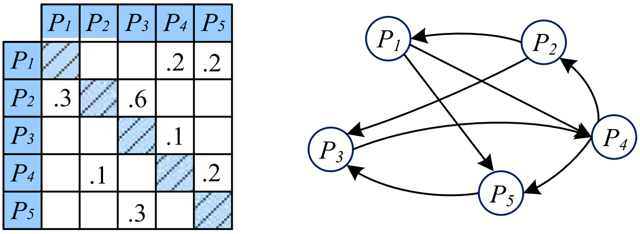
Review sentiment reflects customer satisfaction. Customers express positive sentiment in reviews to show their satisfaction with products. Contrariwise, negative sentiment reviews indicate dissatisfaction from customers [45]. Products with more positive sentiment reviews yield better performance. Therefore, we use sentiment analysis to estimate product performance. Given that sentiments are affected by many product-unrelated factors such as consensus, we develop a filter to exclude feature-unrelated reviews. For each review, the filter decides whether the review contains feature-related words provided by LDA. If the review has words (or synonyms) offered by LDA, the review remains for analysis. Otherwise, the review is excluded from our datasets.

We use machine-learning methods including Random Forest (RF), k-Nearest Neighbor (KNN), Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM) and AdaBoost (Ada) to identify the sentiment of sentences. Machine-learning methods treat the task of sentiment analysis as text classification [49]. The first step of machine-learning based sentiment analysis is to transform unstructured texts into numerical vectors. Then these vectors are used as input of machine-learning methods for text sentiment classification [50]. We use the famous *Doc2Vec* model [51] to transfer texts into numerical vectors and used labeled data to train machine learning models. Then we utilize trained machine learning models to discover sentences' sentiments.

Based on sentence sentiments, we calculate the feature sentiment scores which are obtained by the difference between the number of positive sentences and the number of negative sentences. Then feature sentiment scores are averaged by the number of sentences. Feature sentiment scores indicate the feature performance. The higher the scores, the better the feature performance.

3.3 Comparative network construction

With feature sentiment scores, we discover the comparative relations among products and construct comparative networks. The comparative relations indicate whether two products are comparable. We use Product Common Feature Score Matrix (PCFSM) and Product Unique Feature Score Matrix (PUFSM) to obtain product performance and comparative relationships. In PCFSM (PUFSM), rows denote products and columns denote common (unique) feature sentiment scores. We define values of PCFSM (PUFSM) as PF_{ij} and $i=1, 2,\ldots, m$, $j=1, 2,\ldots, l$. l is the number of common (unique) features. Common features are mentioned by customers in most products' reviews, which means most customers care about the performance of common features when evaluating product performance. For a certain product, its good common feature performance will not satisfy customers a lot, but the poor common features performance will induce great dissatisfaction of customers. On the other hand, unique features usually are impressive features that attract customers. Unique features with good performance can surprise



(a) An example of CRM (b) An example of CRN

Fig. 1 Examples of CRM and CRN

customers, while products without unique features will not dissatisfy customers too much if they perform their common features reliably. Therefore, we use the feature differences to estimate comparative relations with the distinguishing of common and unique features. Detailed equations are shown in Eq. (1)–(2),

$$
\begin{aligned}
CScore_{ik} &= \sum_{j=1}^{l} \left(w_i \cdot PF_{ij} - w_k \cdot PF_{kj}\right), \\
CR^{c}_{ik} &= \begin{cases} W \cdot CScore_{ik}, CScore_{ik} \le 0 \\ CScore_{ik}, CScore_{ik} > 0 \end{cases}, \\
CR^{u}_{ik} &= \begin{cases} 0, CScore_{ik} \le 0 \\ W \cdot CScore_{ik}, CScore_{ik} > 0 \end{cases},
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
normalized\_CR^{c}_{ik} &= \frac{CR^{c}_{ik} - \min(CRM_c)}{\max(CRM_c) - \min(CRM_c)} \\
normalized\_CR^{u}_{ik} &= \frac{CR^{u}_{ik}}{\max(CRM_u)}
\end{aligned}
\tag{2}
$$

where w_i and w_k are the weights of product features. The weights can be obtained by expert evaluation through Analytic Hierarchy Process (AHP). For brevity, the details of weight measurement via AHP are introduced in Appendix A. In Eq. (1), $CScore_{ik}$ is the aggregate difference between two products on all feature performance. $CScore_{ik} > 0$ means that product P_i outperforms P_k while $CScore_{ik} < 0$ means P_k outperforms P_i. CR^{c}_{ik} (CR^{u}_{ik}) is the comparative relation between P_i and P_k which is derived from common (unique) features. W is the parameter to amplify differences of features. Given that relation strengths are over 0, we use Eq. (2) to make CR^{c}_{ik} (CR^{u}_{ik}) range from 0 to 1 and obtain $normalized\_CR^{c}_{ik}$ ($normalized\_CR^{u}_{ik}$). $normalized\_CR^{c}_{ik}$ ($normalized\_CR^{u}_{ik}$) represents that to what extent P_i is comparable with P_k. The larger $normalized\_CR^{c}_{ik}$ ($normalized\_CR^{u}_{ik}$) indicates P_i is more comparable P_k. We use Comparative Relations Matrix (CRM) to represent these normalized CR values. Taking P_2 and P_3 in Fig. 1(a) for example, $normalized\_CR_{23} = 0.6$ but $normalized\_CR_{32} = 0$, indicating P_2 is superior to P_3 with a value of 0.6. But when we focus on P_3, $normalized\_CR_{32}$ valued 0 means it poses no threat to P_2.

With Eq. (1)–(2), PCSFM (PUFSM) is transferred into CRM based common (unique) features. We define CRM based common features (unique features) as CRM_c (CRM_u). The total CRM is calculated by $\mathrm{CRM}_c + \mathrm{CRM}_u$. Based on the total CRM, product comparative networks are constructed. In this work, we use Comparative Relation Network (CRN) to display comparative relations. Figure 1(b) shows the CRN related to CRM in Fig. 1(a).

CRM reflects the direct comparative relationships among products. However, there may exist indirect comparative relations between two products. Observing Fig. 1, P_4 doesn't have a comparative relation with P_3 ($normalized\_CR_{43}=0$). But in CRM, P_4 is predominant to P_5 with a value of 0.2 and P_5 is predominant to P_3 with a value of 0.3. So P_4 has an indirect comparative relation to P_3 ($P_4 \rightarrow P_5 \rightarrow P_3$).

To measure direct and indirect comparative relations, we define Integrated Comparability (IC) to measure accurate relationships by considering influence propagating within three steps. IC_{ij} means that P_i has an integrated comparative relation with P_j. When considering the propagation paths with 1 step, the IC_{ij} is $normalized\_CR_{ij}$ which indicates the direct comparative relation between P_i and P_j. When considering 2-step propagation paths, IC_{ij} equals the sum of the direct relation between P_i and P_j and all influences propagating from P_i to P_j by passing one product. Similarly, for the propagation paths with 3 steps. Besides, the influence will decrease with the extension of propagation paths. Hence, the comparability will decrease as the path extends. Hence, IC_{ij} is calculated through Eq. (3),

$$\begin{aligned} IC_{ij} = normalized\_CR_{ij} &+ \sum_{h=1}^{m} (normalized\_CR_{ih} + \lambda_I \cdot normalized\_CR_{hj}) \\ &+ \sum_{h=1}^{m} \sum_{k=1}^{m} (normalized\_CR_{ih} + \lambda_I \cdot normalized\_CR_{hk} + \lambda_{II} \cdot normalized\_CR_{kj}), \end{aligned} \tag{3}$$

where λ_I and λ_{II} are decay parameters and $0<\lambda_{II}<\lambda_I<1$. After normalizing IC, we construct the Integrated Comparativeness Matrix (ICM) and the Integrated Comparativeness Network (ICN). In the calculation of CR and IC, we consider that products do not compare with themselves. Thereby the values of the diagonal line in CRM and ICM are all 0.

3.4 *Competitive analysis* via *ID-KS*

After revealing comparative relationships among products, we conduct a competitive analysis via the proposed ID-KS. Most studies of the network-based competitive analysis use network partition methods in the area of SNA mentioned in Sect. 2.3 (we call them classical methods). But these classical methods have several deficiencies. Firstly, when most node degrees are very close, classical methods will make the network partition ineffectively. Nodes with different importance will be classified into the same class (see node a and b in Fig. 2, node b connects with two nodes and influences more nodes than node a, but they are assigned the same KS value in classical k-shell decomposition). Secondly, classical methods are developed for undirected graphs. They cannot deal with the weighted and directed network like

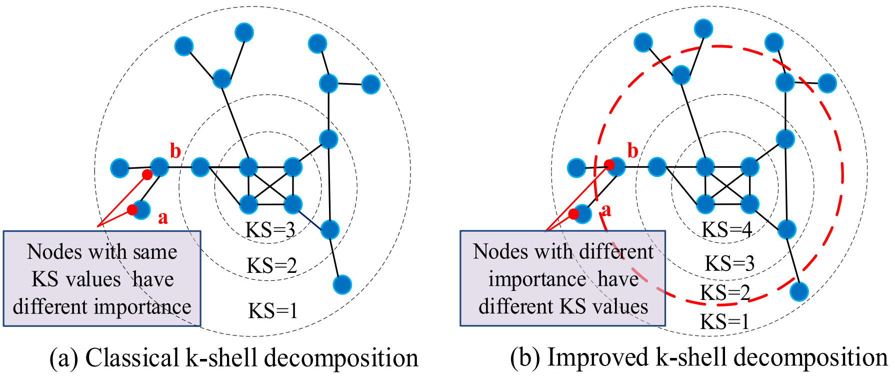


Fig. 2 Classical and improved k-shell decomposition

ICN. The third weakness is that classical methods don't take node attributes into account. This drawback results in insufficient exploitation of information. Our proposed ID-KS tackles these disadvantages above and partition directed networks more effectively. It can attain more objectives including competitor identification, product comparison, product ranking, brand comparison and market structure analysis than extant studies.

ID-KS includes two parts. The first part is to calculate the degree for each node. Node degree is the basis of the network partition. In the directed networks, node degree can be divided into in-degree d_I and out-degree d_O. A node's in-degree is the number of edges that point to the node while out-degree is the number of links to other nodes. In other words, in-degree is the impact exerted by the node's neighbors and out-degree is the impact that a node has on its neighbors. For a certain node, if its out-degree is larger than in-degree, the node is influential to other nodes. Otherwise, the node is affected by other nodes. To take a thorough consideration of in-degree, out-degree and node attributes, we develop a special node degree named integrated degree (ID) and define the integrated degree of product P_i as $ID(P_i)$. $ID(P_i)$ equals the difference between P_i's out-degree and in-degree and is calculated by Eq. (4)–(5).

$$ID(P_i) = d_O(P_i)^{\gamma} \cdot (\sum_{k=1}^{m} VF(P_i) \cdot IC_{ik})^{\eta} - d_I(P_i)^{\gamma} \cdot (\sum_{k=1}^{m} VF(P_k) \cdot IC_{ki})^{\eta}, \tag{4}$$

$$VF(P_i) = \sum_{t=1}^{T} \frac{vf_t(P_i)}{\max(\mathbf{vf_t})}. \tag{5}$$

In Eq. (4), the left term denotes the product of out-degree and node attributes. $\sum_{k=1}^{m} VF(P_i) \cdot IC_{ik}$ indicates the aggregated influence of P_i 's node attributes. For P_i, the node attribute influence that it has on other products is measured by the product of node attribute values and the integrated comparability relations that P_i has with other products. γ and η are adjusted parameters that need to satisfy

Table 2 The algorithm of the improved network decomposition

Algorithm 1: Improved network decomposition

Input: Integrated competitiveness network, Refinement parameter α
Output: KS values for each node
Steps:

i=1
repeat
  for each node in the network **do**
    Calculate integrated degree $ID(P_i)$ by Eq (3)-(4)
    Normalize integrated degree by $ID(P_i)/|min(ID(P_i))|$
    if existing product P_i with negative $ID(P_i)$
      for each P_i with negative $ID(P_i)$
        Remove nodes with $ID(P_i) \in [-1, \alpha - 1]$, α is the refinement parameter and $\alpha \in [0,1)$
      end for
    else
      for each P_i
        Remove nodes with $ID(P_i) \in [1, 1 + \alpha]$
      end for
    end if
    KS value of removed nodes equals to i
  end for
until all nodes have KS values

constraints of $0 \le \gamma, \eta \le 1$ and $\gamma + \eta = 1$. $VF(P_i)$ denotes the integrated attribute values of P_i and it is derived by Eq. (5). $vf_t(P_i)$ denotes the tth attribute value of P_i and $\mathbf{vf_t}$ is the tth attribute value vector of all products. T is the number of node attributes. The integrated degree is different from the traditional degree because it may be negative. Products with negative integrated degrees receive threats from other products and cannot compete with others. In contrast, the positive integrated degrees indicate products are more competitive.

The second part of ID-KS is to decompose networks. To overcome the shortcomings of classical methods, we improve the partition process of k-shell decomposition and detail it in the algorithm in Table 2.

After the decomposition process, each node has a KS value and ICN is divided into several clusters based on KS values. Then we make a competitive analysis with the results of ID-KS.

(1) *Product comparison*. We compare products from the perspective of products and features. From the perspective of products, products with greater KS values are more competitive than those with less KS values. From the perspective of features, we compare the feature performance of different products by PCFSM and PUFSM. For a certain product, dissatisfied features with negative senti-

Table 3 Data description

	Car dataset	Laptop dataset
Number of products	24	20
Number of brands	9	9
Number of reviews	21,743	32,049
Number of reviews with comparative sentences	760 (3.50%)	716 (2.23%)
Average length of reviews (words)	154.88	48.79
Minimum length of reviews (words)	1	1
Maximum length of reviews (words)	1058	332
Proportion of short/medium/long reviews*	16%/19%/65%	14%/51%/35%

*Criteria of short/medium/long English reviews is length ≤ 50/50 < length ≤ 100/length > 100. Criteria of short/medium/long Chinese reviews is length ≤ 20/20 < length ≤ 50/length > 50

ment scores are significant weaknesses of the product. Besides, in contrast to its competitors, this product is also weak on features with smaller feature sentiment scores. Manufacturers should pay attention to these product weaknesses for product improvement.

(2) *Competitor identification*. We identify competitors according to the following guidelines: (i) Products owning the same KS value are mutual competitors. (ii) Products with maximum KS value are top competitors in the market.

(3) *Product ranking*. We rank products by setting α as 0. After decomposing using ID-KS, each product has a unique KS value. The higher the KS values, the higher the product rankings.

(4) *Brand comparison*. Brand comparison is conducted by drawing a parallel between products belonging to assorted brands. We measure the competitiveness of a brand by calculating the average KS values of its products. Brands with higher competitiveness are more attractive than those owning less competitiveness.

(5) *Market structure analysis*. Market structure analysis is to explore the complementary relationships among the products or brands [52]. When the market structure analysis can be conducted in an automated and prompt way, marketing research and managerial decisions can be made more precise [53]. In our study, ID-KS clusters products into several categories when α is larger than 0. Products in the cluster with larger KS values stay in an advantageous position in the market while products with smaller KS values are in the adverse place of the market.

Table 4 Product common features and unique features

Features	Car dataset	Laptop dataset
Common features	Seat, driving experience, fuel consumption, driving performance, dealer service, space, reliability	Screen, customer service, running speed, startup speed, appearance, office work performance
Unique features	Interior, assistant system, transmission, appearance, control, power, quality, steering wheel, console, sunroof, entertainment, cost performance	Heat dissipation, cost performance, logistics, weight, keyboard, configuration, game performance, price, workmanship, system, complimentary products, after-sales, accessories, quality, office software, H-Share, packaging, easy to use, audio

4 Case study

4.1 Dataset

We use two cases, SUV cars and laptops, to illustrate the effectiveness and generality of our method. We choose the popular car review website named Edmunds[1] and the Chinese online platform, JD.com,[2] to collect online reviews. The detailed data description is shown in Table 3.

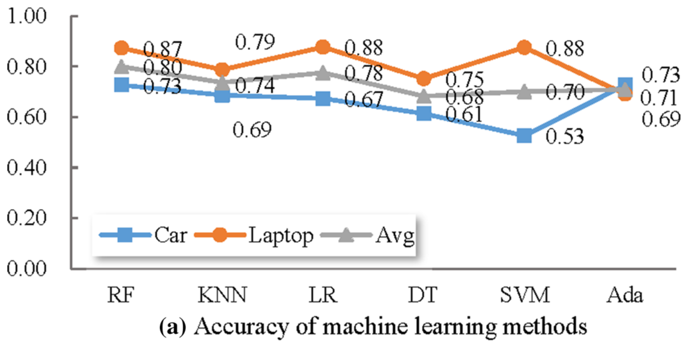
For the car dataset, we use P1 ~ P24 to represent the 24 car models and PB1 ~ PB9 to represent 9 brands. In addition, we choose the number of reviews, average rate and average *Agreed rate* as node attributes, and present them in Appendix B. *Agreed rate* is the percentage customers who feel the reviews to be helpful, and it is obtained by Eq. (6)

$$
Agreed\,rate = \begin{cases} \sqrt{\dfrac{helpful \times vote}{(vote - helpful) + 1}}, helpful \geq \dfrac{vote}{2} > 0 \\ \sqrt{\dfrac{helpful}{(vote - helpful) \times vote}}, 0 \leq helpful < \dfrac{vote}{2} \\ 0, helpful = vote = 0 \end{cases} \tag{6}
$$

where *helpful* is the number of customers who feel the focal review is helpful, and *vote* is the number of customers who vote on whether the focal review is helpful. *Agreed rate* indicates how helpful the review is. The larger the *Agreed rate*, the more customers agree that the review is useful. After averaging all reviews' *Agreed rate* for a certain product, the average *Agreed rate* is used as the node attribute of the product. For the laptop dataset, we use C1 ~ C20 to represent 20 laptops and

[1] https://www.edmunds.com.

[2] https://www.jd.com.



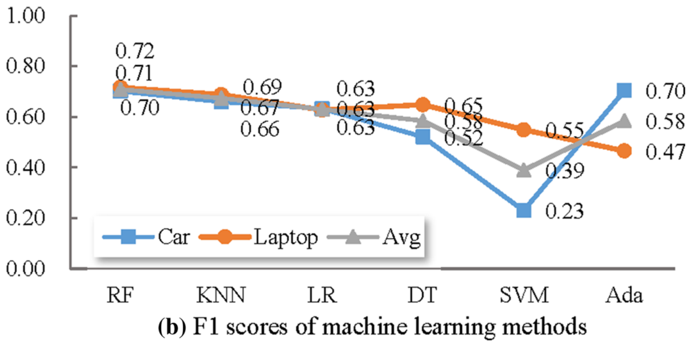
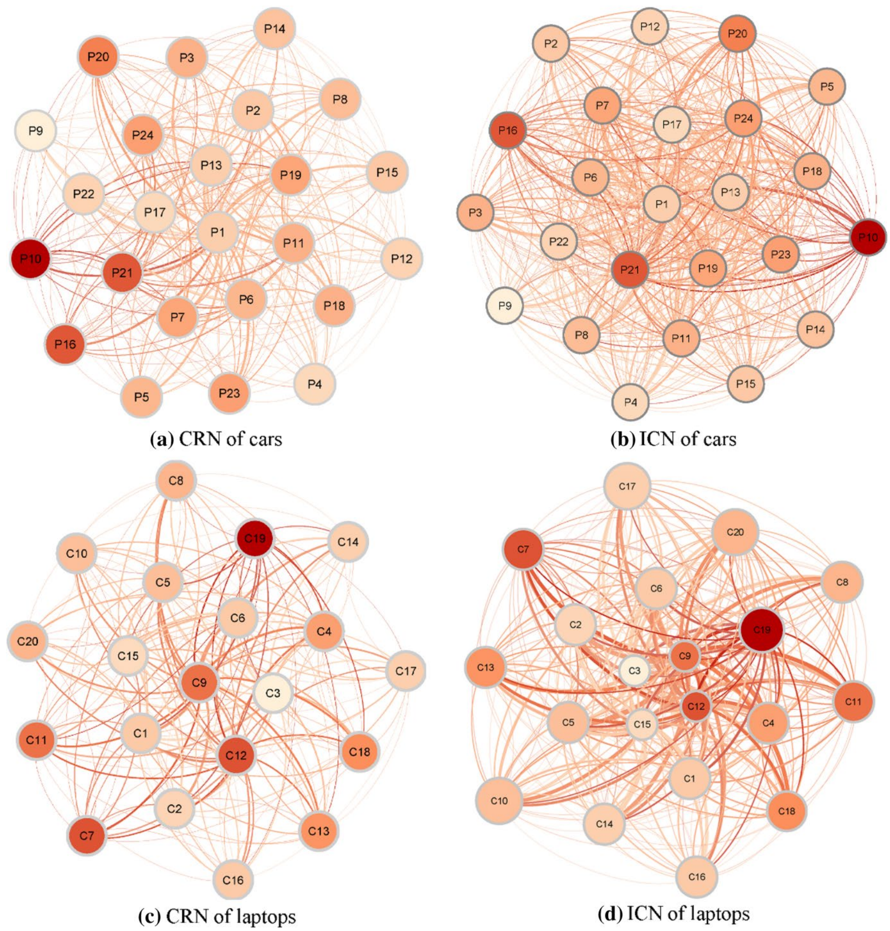


Fig. 3 Performance of sentiment classification via different machine learning methods

LB1 ~ LB9 to stand for 9 brands. Node attributes used in the laptop case are the percentage of five-star rate (5SR), total useful votes (TUV) and average length per review (measured by the number of words) (ALPR), and are shown in Appendix C. In the phase of LDA, we choose the number of topics to be 20 for cars and 15 for laptops. It is noteworthy that the review lengths of both car and laptop datasets shown in Table 2 are balanced. Most reviews are medium or long reviews, which prevent LDA from suffering from short texts. After filtering the duplicated features, we get detailed topics (displayed in Table 4) for two datasets. In the process of sentence filtering, English synonyms are provided by a famous lexicon named WordNet, while Chinese synonyms are offered through the lexicon built by Wang and Hu.[3]

In sentiment analysis, we select six effective machine learning methods aforementioned in Sect. 3.2. The training dataset for sentiment analysis is composed of 2000 reviews with sentiment labels. We get these sentiment labels by manual reading and tagging. Three undergraduate students majoring in Industrial Engineering at Tianjin University tagged these 2000 reviews. Each review was tagged by the three students. The inter-annotator agreement rate of the manual tagging is 90.13%. In the process of data training, we use 10-cross validation to train these models, and we

[3] https://github.com/chatopera/Synonyms.



(a) CRN of cars

(b) ICN of cars

(c) CRN of laptops

(d) ICN of laptops

Fig. 4 CRNs and ICNs

present the accuracy and F1 scores in Fig. 3. From Fig. 3 we see that RF has the best performance in sentiment classification. Hence, we use RF to determine the review sentiment.

In the calculating process, parameters are set as follows: $W=2$, $\alpha=0.5$, $\lambda_{\mathrm{I}}=0.7$, $\lambda_{\mathrm{II}}=0.3$, $\gamma=\eta=0.5$. H values are 20 for the car dataset and 15 for the laptop dataset. For brevity, we don't report the CRMs and ICMs. Instead, we use CRNs and ICNs in Fig. 4 to derive competitive insights. The size and color of nodes in both networks indicate degree (here we use the sum of in-degree and out-degree as node degree) and node attribute values. The larger the node size, the larger the node degree. And the darker the color, the larger the node attribute values. All the calculation processes are done using Python 3.7.

Table 5 Product rankings

Dataset	Ranking
Cars	P1, P6, P17, P11, P14, P24, P3, P5, P2, P12, P10, P15, P16, P20, P23, P4, P18, P21, P8, P7, P9, P22, P19, P13
Laptops	C9, C3, C15, C12, C19, C17, C16, C7, C18, C20, C10, C11, C14, C8, C13, C5, C2, C4, C6, C1

Table 6 ID-KS results of cars

KS	Car models	Average KS	Car brands
6	P1	4	PB1, PB6
5	P6	3.2	PB4
4	P3, P11, P14, P17, P24	3	PB2
3	P2, P5, P10, P12, P15, P16	2.5	PB7
2	P4, P7, P8, P18, P20, P21, P23	2	PB3, PB5, PB9
1	P9, P13,P19, P22	1.5	PB8

4.2 Competitive analysis of SUV cars

Car performance comparison. We first analyze the feature performance of each car model using PFSM in Appendix D1 and D2. The bold values in Appendix D1 and D2 denote the best score for each product features or the negative scores indicating poor feature performance. PFSM reveals product performance on common features and unique features. We bold the value for each column to display which product has the best performance in related features. As we can see from Appendix D1 and D2, when focusing on common features, P21 performs best in fuel consumption and driving performance, while P2 outperforms others in the seat and space. P19 shows its competitiveness in driving experience. P22 achieves good dealer service. P20 shows good performance in reliability. However, P17 gets a negative score in fuel consumption, which means P17 has the highest fuel consumption among 24 cars. Manufacturers of P17 should make efforts to reduce fuel consumption. From the perspective of unique features, most customers mentioned the features of the interior, assistant system and transmission. P14, P11 and P3 are the top cars in these unique features. These cars with good performance on unique features will arouse close interest from customers. As for product comparison, we compare any two cars by comparing their feature sentiment scores. Cars with higher feature sentiment scores own better performance.

Car ranking. We obtain the ranking of cars by setting α to 0. The ranking results are presented in Table 5. Among 24 cars, P1, P6, P17 are the top 3 cars, while P22, P19 and P13 are the bottom 3 cars.

Competitor identification for cars. After comparing car performances, we identify competitors based on CRN, ICN and ID-KS. CRN and ICN describe the

Table 7 ID-KS results of laptops

KS	Laptops	Average KS	Laptop brands
6	C9	3.5	LB5
5	C3	3	LB7
4	C12, C15	2.75	LB1
3	C7, C16, C17, C19	2.5	LB8, LB9
2	C8, C10, C11, C13, C14, C18, C20	2.3	LB6
1	C1, C2, C4, C5, C6	2	LB2
–	–	1	LB3, LB4

comparative relations derived from PFSM. In Fig. 4, the number of relations exhibited in ICN is more than that of CRN. This discrepancy indicates the correction of our methods to consider indirect relations. We then identify competitors based on the results of ID-KS in Table 6. Among these 24 cars, P1 is the most competitive car and the top competitor in the market. Cars with the same KS values are competitors. Taking P5 for example, the competitors of P5 are P2, P10, P12, P15 and P16, as they are all assigned with the same KS value 3. When comparing with its competitor like P12, P5 is inferior in the common aspects of seat and driving performance. It is weak in the unique features of reliability and control. Hence, the manufacturer of P5 should improve the inferior performance and gain competitive advantages by narrowing the gaps between P5 and its competitors.

Competitive analysis among car brands. To analyze the competitive relations among nine car brands, we calculate the average KS values for each brand and show them in Table 6. From Table 6 we observe that PB1 and PB6 are the most dominant brands. Their products have overwhelming competitiveness in the market. Brand PB8 gets into a stressful situation because it cannot compete with other brands. It is urgent for PB8's manufacturer to improve its products and gain competitive edge. Some brands are less competitive due to significant differences in their cars. Taking PB7 for example, its three models (P22, P23, P24) have different competitiveness. P24 has the KS value of 4 while P22 owns the KS value of 1. Manufacturers of PB7 should find the gaps between these car models and upgrade P22 comprehensively.

Market structure analysis for cars. From the result of ID-KS, we find that 24 cars are classified into 6 clusters. The performance of P1 and P6 is overwhelming in the market. These two products have absolute competitive edges in contrast to the remaining cars. The remaining four clusters contain a similar number of cars, which means a significant performance gap exists in the current SUV market. Those vehicles with worse performance should strive to improve quality and narrow the gap with high-performance vehicles.

4.3 Competitive analysis of laptops

Laptop performance comparison. PFSM of 20 laptops in Appendix E1 and E2 provides information about product performance in 25 product features. Similarly, the bold values in Appendix E1 and E2 show the best score for each product features or

Table 8 Performance of ID-KS in competitor identification and brand comparison

	The car dataset	The laptop dataset
Accuracy of competitor identification	0.8750	0.8500
Accuracy of brand comparison	0.8667	0.5500

the negative scores indicating poor feature performance. Among these laptops, C15 has an excellent performance in screen and startup speed. C12 attracts customers through its excellent performance on running speed and appearance. C11 is appreciated for its customer service quality but C16 dissatisfies its customers for the worst service. C1 gets negative scores in the aspects of free gifts and aftersales which means the manufacturer of C1 encounters after-sales problems for free gifts. For any two laptops, we compare their performance according to their feature sentiment scores. Laptops with higher scores own better performance.

Competitor identification and product ranking. According to the results of ID-KS in Table 7, we discover competitors for each laptop and rank these laptops (Ranking results are shown in Table 5). Laptops with the same KS values are competitors. C9 is the top competitor among 20 laptops. It also wins first place in the laptop ranking. In contrast, C1, C2, C4, C5 and C6 have the smallest KS value and are competitors. But they rank bottom for poor performance. More efforts are needed for these three laptops to enhance their competitiveness.

Market structure analysis and brand comparison. 20 laptops are categorized into 6 clusters and most laptops are classified into clusters with small KS values. This pyramid structure means most laptops in the market do not have outstanding performance. Manufacturers should improve their products and promote the development of the laptop market. In the aspect of brand comparison, the average KS values in Table 7 show that LB5 is the predominant brand in the market. Though LB2 has 3 laptop series, it only locates in fifth place due to the poor performance of C2 and C4. LB3 and LB4 are the worst brands among 9 brands. Their manufacturers need to expand their influence and enhance word-of-mouth by making a great improvement in product performance.

4.4 Effectiveness validation of ID-KS

Accuracy of competitor identification and brand comparison. To validate the accuracy of ID-KS in competitor identification, we gather competitor information from famous third-party comparison websites for both car and laptop datasets. For the car dataset, we discover the competitor for each car from Edmunds, which provides the information of the most concerned car except the focal car. For the laptop dataset, we search competitors from ZOL.com (www.zol.com), which is a famous three-party Chinese comparison website. With the competitor information, we measure the accuracy of ID-KS in competitor identification by estimating the percentage of products whose competitors have been predicted correctly. In addition, we compare the brand comparison results with the word-of-mouth ranking from the websites

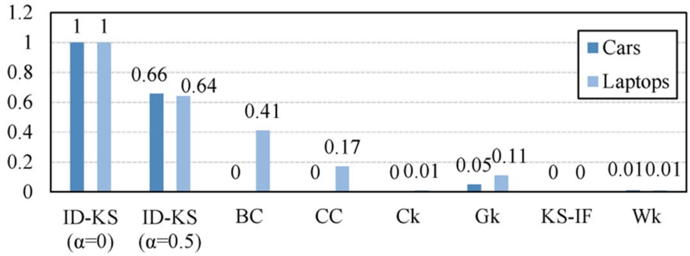


Fig. 5 Monotonicity results of ID-KS and benchmarks

above. In this comparison, we set α equals 0 and obtain the ranking of brands. Then Spearman Rank Correlation [11] is used for the accuracy measurement of brand ranking. Both results are displayed in Table 8. From Table 8 we observe that ID-KS achieves good performance in competitor discovery. Nearly 85% of products are predicted correctly. The accurate result of competitor discovery also suggests that the market structure derived by ID-KS is reasonable, because we identify competitors based on the results of the market structure. In the aspect of the brand ranking, our results are closed to the word of mouth results, indicating that our method is effective in brand comparison and reflects the real word of mouth of products.

Effectiveness of network partition. To validate the effectiveness of ID-KS in network partition, we select the classical k-shell decomposition (Ck) [32] and its improved approaches including K-shell iteration factor (KS-IF) [33], Weighted k-shell by Garas (Gk) [34] and Weighted k-shell by Wei (Wk) [35] as benchmarks. We also compare the degree method and the centrality methods including Closeness Centrality (CC) and Betweenness Centrality (BC) [31].

Approaches with an outstanding capability to differentiate nodes will assign nodes with different KS values (centrality values for centrality methods). To evaluate the effectiveness of different methods, we adopt the evaluation method named monotonicity [38]. Monotonicity provides a score between 0 and 1. If every node in the network has a unique KS value (or centrality value), then the monotonicity equals 1. Whereas, the monotonicity is less than 1. Methods with larger monotonicity values are better than others. Monotonicity can be calculated by Eq. (7).

$$M(R) = \left(1 - \frac{\sum_{r\in R} n_r(n_r - 1)}{n(n-1)}\right)^2. \tag{7}$$

In Eq. (7), r is the rank value which is the KS value (centrality value) in our research. R denotes the rank vector. Numerical values in R mean the number of nodes for each rank value. n_r means the number of nodes with rank value r. n is the total number of nodes in the network.

The results of contrast experiments are shown in Fig. 5. From Fig. 5, we observe a clear outperformance of ID-KS. Betweenness centrality, closeness centrality and degree centrality can only roughly distinguish nodes because most of the nodes have

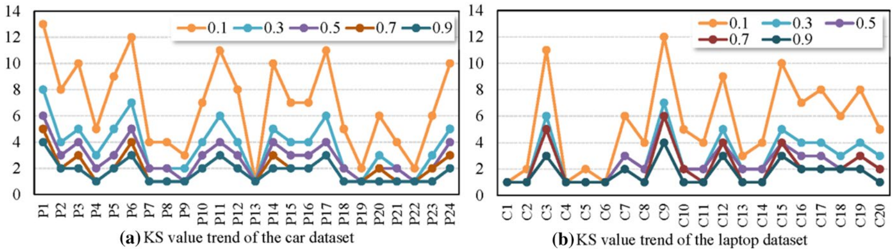


Fig. 6 The influence of α on model results

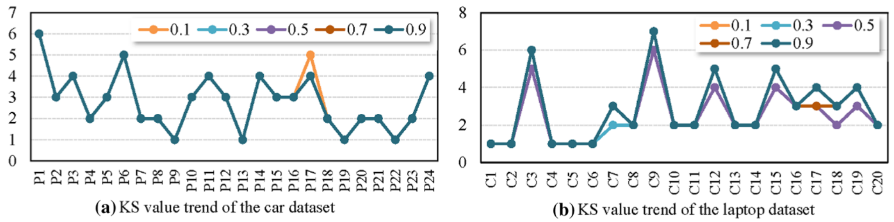


Fig. 7 The influence of γ (η) on model results

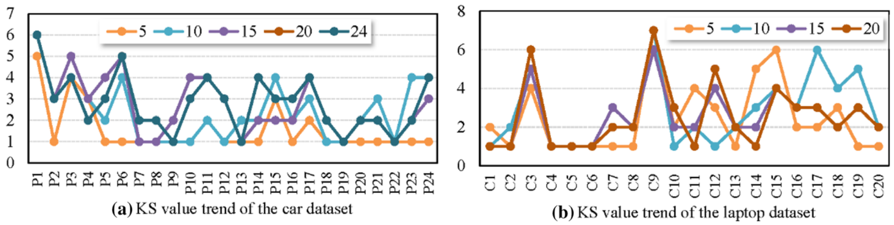


Fig. 8 The influence of H on model results

the same centrality value. Figure 5 also shows that our method outperforms classical k-shell decomposition and its improved methods. Thus, ID-KS stands out from baseline methods and is more effective in the network partition.

4.5 Parameter analysis

4.5.1 Sensitive analysis of the single parameter

In ID-KS, five parameters (α, γ, η, H and W) affect the decomposition results. Refinement parameter α is closely related to the network decomposition and affects KS values to a considerable extent. Figure 6 summarizes the marked variation of KS values for each product when α increases. Different broken lines have a similar variation in KS values, but the gap between two broken lines is narrowed if their related

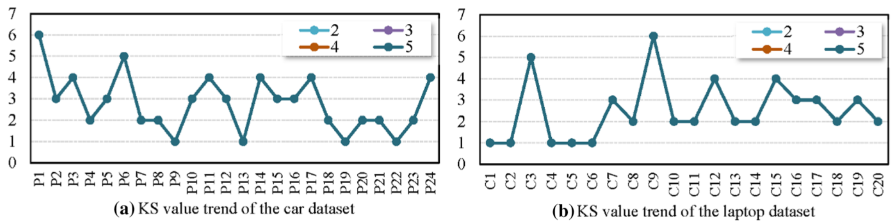


Fig. 9 The influence of W on model results

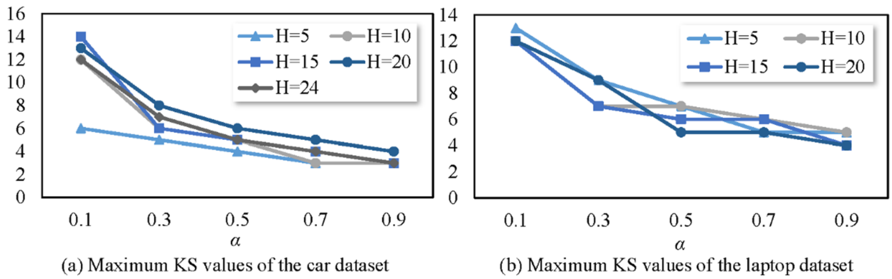


Fig. 10 The influence of α and H on model results

α values are close. Therefore, the distinction in KS values between two nodes will enlarge when α decreases.

Figure 7 depicts adjusted parameters γ and η's effects on decomposition results. All figures in Fig. 7 express that γ and η have slight impacts on partition results. The trend of KS values has only minimal fluctuations. Therefore, γ and η have negligible influences on partition results.

Except for α, γ and η, we explore the influences of H and W on model results. Figure 8 illustrates the KS value trend when H changes. From Fig. 8 we find that KS values have considerable fluctuations with the increase of H. When H rises from 5 to 10, most products' KS values change. But when H increases from 15 to 24 (20 for the laptop dataset), the lines of KS values are overlapping which means that most products have no change in their KS values. Thereby, we conclude that H has an impact on the final results. When H is small, the impact is great. But when H increases, the impact becomes marginal. Figure 9 displays W's impacts on ID-KS results, which shows that W has no impact on model results.

4.5.2 Sensitivity analysis of multiple parameters

Given that multi-parameter combinations may have different effects on final results, we conduct the sensitivity analysis with the consideration of multiple parameters. Based on the results in Sect. 4.5.1, we exclude the parameters γ, η and W in the multi-parameter sensitivity analysis because they have negligible influences on model results. Hence, we research the impact of α and H on ID-KS results by studying the trend of maximum KS values. The underlying reason for studying maximum

KS values is that maximum KS values indicate how many clusters that products are categorized into. The number of clusters describes the change of KS values indirectly. When the maximum KS value increases, products are classified into more categories, and the KS value of the product changes greatly. Figure 10 presents detailed results of two datasets. From both two figures in Fig. 10, we observe an obvious decrease in the maximum of KS values. This trend denotes that KS values fall markedly with an increase of α and is consistent with the results in Fig. 6. As for H, the maximum KS values fluctuate with the change of H, but the results of different H get close especially when H is over 15. From the perspective of parameter selection, we find that the maximum KS values have minimal change when $\alpha \geq 0.5$. When α equals 0.7 or 0.9, the number of clusters is small and products cannot be distinguished clearly. When α equals 0.1 or 0.3, the number of clusters is large and the market structure cannot be identified clearly. Moreover, the trend of maximum KS values also has no change when H is bigger than 20 (15 for the laptop dataset). Therefore, the parameter selection ($\alpha = 0.5$, $H = 20$ (15 for the laptop dataset)) in the case study is reasonable.

To summarize the results of parameter analysis, we find that refinement parameter α has the greatest impact on decomposition results. When α rises, the gaps between nodes will decrease gradually. H makes ID-KS results fluctuate especially when H is small. When H gets close to the number of products, H's influence becomes minimal. γ, η and W have a negligible influence on decomposition results. The changes in these three parameters will not lead to the obvious variation in the KS values.

5 Theoretical and managerial implications

5.1 Theoretical implications

The utilization of online reviews has become a surge in the extant studies of competitive analysis. These studies conduct product ranking, product comparison or competitor identification via deriving customer opinions from online reviews. Compared with previous literature, this study focuses on product ranking, product comparison or competitor identification simultaneously instead of only studying one research problem. Moreover, more competitive information on market structure and brand comparison is provided. In conclusion, our study has three main theoretical implications and contributions.

First, our study provides more competitive information, including product comparison, product ranking, competitor identification, market structure analysis and brand comparison, than existing literature. The proposed method in our study conducts competitive analysis via online review more comprehensively and effectively.

Second, in our competitive analysis framework, we differentiate common and unique features when evaluating product performance based on the performance of features. The distinguishment of common and unique features makes competitive analysis more reasonable and comprehensive. Besides, when estimating comparative relations via comparative networks, we consider the influence propagation which helps us discover indirect comparative relations. This consideration also enhances the comprehensiveness and accuracy of our results.

Third, when decomposing comparative networks, we develop the improved k-shell approach named ID-KS which can be utilized in direct networks. The experimental results validate the effectiveness of ID-KS in network partition.

5.2 Managerial implications

Competitive analysis is essential for manufacturers to improve their products and gain competitive edge. With the popularity of online reviews in competitive analysis, more and more manufacturers have to glean competitive information from online reviews. But the colossal volume of unstructured texts makes them bewildered when facing online reviews. Our study aims to help manufacturers deal with online reviews more automatically and effectively in the process of competitive analysis. The managerial implications of our approach can be summarized into three folds.

First, our study assists manufacturers to identify competitors. Identifying competitors correctly is critical for product comparison and competitive analysis. Based on the comparative networks, our study finds out competitors for a certain product by applying the improved SNA method in comparative networks. With the identified competitors, manufacturers can further compare products and uncover product weaknesses accurately.

Second, our method helps manufacturers discover product weaknesses and improve directions via product comparison. In our research framework, two PFSMs, that is PCFSM and PUFSM, provide the detailed feature performance comparison among products. For a certain product P_i, its manufacturer can compare it with its competitor to find which features are less competitive. Manufacturers can find performance gaps and improve their products by finding which competitors perform better than P_i in these uncompetitive features. For different features, manufacturers need to take different improvement actions. If P_i is uncompetitive in common features, the manufacturer should find the reasons for the failure of common features and enhance common feature performance at once. If P_i does not own common features that most similar products in the market have, the manufacturer should consider adding these common features when redesigning P_i or developing new products. If P_i does not have or performs badly in unique features, the manufacturer can enhance unique feature performance or supplement unique features after P_i ‘s common features have been improved.

Third, the information on market structure and brand comparison helps manufacturers devise their development strategies. With brand comparison information, manufacturers can have a comprehensive understanding of the leading brand in the market. According to the market structure information, the manufacturer can find the performance gaps among its different products. This information helps manufacturers make better development strategies for their brand and various products.

6 Conclusion

In this study, we propose an innovative method named ID-KS to implement a competitive analysis. Given the differences between product common and unique features, we use LDA and sentiment analysis to estimate product performance and construct a comparative network ICN which measures direct and indirect comparative relations. We use ID-KS to derive competitive information from ICN which overcomes the drawback of classical SNA methods. ID-KS helps us conduct competitive analysis including product comparison, product ranking, competitor identification, brand comparison and market structure analysis. Finally, we use case studies to validate the performance of ID-KS.

There are some limitations in our research that can be extended in the future. First, when using LDA, we choose the number of topics with subjectivity that is similar to most research. How to select a reasonable topic number needs further research. Second, we use online reviews and some product attribute data to conduct competitive analysis. Social features of reviewers can be further considered to weight different reviews and make the analysis more reliable.

Table 9 Meanings of nine-scale scores (evaluate PF_2 comparing to PF_1)

Scores	Meanings
1	Indicates two PFs are equally important
3	Indicates PF_2 is moderately more important than PF_1
5	Indicates PF_2 is strongly more important than PF_1
7	Indicates PF_2 is very strongly more important than PF_1
9	Indicates PF_2 is extremely more important than PF_1
2/4/6/8	Indicate the importance of PF_2 is between the scores that closest to the value

Appendix A: Details of weight measurement via AHP

AHP is a popular approach to estimating the relative importance or weights of assorted factors or decisions. Combining our research objects and the process of AHP proposed by Saaty [54, 55], we illustrate the weight evaluation of product features as follows.

STEP 1: Pairwise comparing product features and estimating their relative importance using the nine-scale scoring. The nine-scale scoring is introduced bellowing table. The comparing results are recorded in the pairwise comparison matrix (PCM). For example, PCM(PF_2, PF_1) denotes the importance of PF_2 comparing to PF_1 (Table 9).

STEP 2: For each PF in PCM, if the PCM(PF_i, PF_j) $= A$, then PCM(PF_j, PF_i) $= 1/A$.

STEP 3: Checking the consistency of PCM by calculating the consistency ratio (CR). Detailed measurement of CR is introduced in the work of Saaty [54]. If CR is smaller than 0.1, PCM is consistent. Otherwise, PCM must be revised until it is consistent.

STEP 4: Normalizing PCM by columns using the sum of the column.
STEP 5: obtaining weights of PFs by averaging the normalized PCM by arrays.

Appendix B: Node attributes of 24 car models

Car	Brand	#Review	Avg.rate	Agreed rate %	Car	Brand	#Review	Avg.rate	Agreed rate %
P1	PB1	854	4.400	77.97	P13	PB5	945	4.043	69.82
P2	PB1	683	4.312	67.45	P14	PB6	420	4.104	77.77
P3	PB2	403	4.304	66.15	P15	PB7	759	4.202	69.00
P4	PB2	316	4.133	70.43	P16	PB4	649	4.336	67.25
P5	PB3	1136	4.053	74.10	P17	PB4	719	4.519	74.02
P6	PB4	799	4.281	74.28	P18	PB8	738	4.547	77.03
P7	PB4	983	4.191	70.68	P19	PB8	1052	3.998	75.82
P8	PB4	572	3.993	72.84	P20	PB9	1203	4.117	71.41
P9	PB3	899	3.904	69.61	P21	PB9	1333	4.278	74.57
P10	PB1	2595	4.344	66.25	P22	PB7	598	4.595	67.77
P11	PB1	1354	4.446	73.95	P23	PB7	1031	4.509	71.95
P12	PB5	433	3.857	74.38	P24	PB7	1275	4.385	69.89

Appendix C: Node attributes of 20 laptops

Laptops	Brand	5SR %	TUV	ALPR	Laptops	Brand	5SR %	TUV	ALPR
C1	LB1	93	1538	8.42	C11	LB6	96	1188	14.58
C2	LB2	94	728	8.33	C12	LB6	98	1612	14.27
C3	LB2	91	699	6.54	C13	LB7	96	1490	17.23
C4	LB2	95	1435	12.25	C14	LB7	97	937	9.83
C5	LB3	94	1119	10.92	C15	LB7	98	865	6.75
C6	LB3	92	1274	10.36	C16	LB7	94	684	13.64
C7	LB4	95	2757	11.44	C17	LB8	94	465	10.06
C8	LB5	95	761	11.10	C18	LB8	94	1419	12.99
C9	LB6	96	1141	15.05	C19	LB9	96	4038	13.37
C10	LB6	96	385	8.74	C20	LB9	97	1405	10.05

Appendix D1: PFSM of car dataset

	Seat	Driving experience	Gas mileage	Driving performance	Dealer service	Space	Reliability	Interior	Assistant system	Transmission	Outlook
w	0.0072	0.0258	0.0530	0.0797	0.0262	0.0071	0.1343	0.0147	0.0170	0.1107	0.0137
P1	0.46	0.56	0.01	0.45	0.36	0.46	1.09	0.86	0.76	0.88	0.69
P2	**0.61**	0.40	0.24	0.40	0.35	0.48	**1.22**	1.04	0.00	0.87	1.14
P3	0.40	0.45	0.41	0.33	0.50	0.47	0.00	0.57	0.00	**1.08**	0.00
P4	0.46	0.37	0.45	0.43	0.32	0.53	0.82	0.00	0.00	0.00	0.72
P5	0.33	0.44	0.42	0.17	0.40	0.43	0.00	0.00	0.69	0.86	0.74
P6	0.34	0.40	0.29	0.35	0.35	0.56	0.64	1.16	0.68	0.78	0.31
P7	0.28	0.23	0.34	0.34	0.38	0.44	0.88	1.06	0.72	0.00	0.00
P8	0.57	0.55	0.39	0.29	0.42	0.45	0.83	1.19	0.00	0.00	0.00
P9	0.33	0.46	0.31	0.36	0.52	0.40	0.63	0.03	0.00	0.00	0.78
P10	0.39	0.33	0.41	0.49	0.43	0.45	0.66	0.00	0.41	0.81	**1.45**
P11	0.50	0.36	0.05	0.28	0.42	0.45	0.88	0.82	**1.06**	0.88	0.87
P12	0.50	0.39	0.42	0.46	0.30	0.31	0.56	0.00	0.77	0.83	0.00
P13	0.46	0.33	0.42	0.38	0.19	0.46	0.93	0.00	0.00	0.00	0.00
P14	0.39	0.22	0.31	0.26	0.41	0.51	0.00	**1.21**	0.90	0.00	0.80
P15	0.43	0.21	0.47	0.21	0.47	0.44	0.00	0.68	0.57	0.00	0.96
P16	0.48	0.20	0.34	0.23	0.48	0.58	1.21	1.10	0.78	0.00	0.00
P17	0.37	0.17	**– 0.10**	0.41	0.50	0.57	1.00	1.21	0.71	0.00	0.88
P18	0.46	0.43	0.31	0.14	0.51	0.45	0.52	0.00	0.76	0.85	0.00
P19	0.36	**0.71**	0.40	0.30	0.44	0.46	0.61	0.00	0.36	0.67	0.00
P20	0.41	0.25	0.25	0.53	0.33	**0.69**	0.94	0.89	0.00	1.04	0.00
P21	0.41	0.13	**0.50**	**0.59**	0.32	0.53	0.00	0.00	0.00	1.00	0.00
P22	0.48	0.08	0.42	0.38	**0.55**	0.33	0.76	1.01	0.00	0.00	0.00
P23	0.46	0.21	0.47	0.44	0.35	0.40	0.66	1.18	0.00	0.00	0.00
P24	0.48	0.35	0.33	0.36	0.48	0.64	0.71	0.90	0.93	0.00	0.00

Appendix D2: PFSM of car dataset

	Control	Power	Quality	Steering wheel	Console	Sunroof	Entertainment	Cost performance
w	0.1318	0.0879	0.1780	0.0262	0.0274	0.0079	0.0102	0.0410
P1	1.00	0.97	1.06	1.00	**1.20**	0.93	0.86	0.74
P2	0.00	0.00	**1.20**	0.00	0.00	0.00	0.00	0.00
P3	0.00	0.76	0.00	1.00	0.00	1.00	0.00	0.00
P4	0.80	0.00	0.00	0.00	0.00	**1.09**	0.00	0.00

	Control	Power	Quality	Steering wheel	Console	Sunroof	Entertainment	Cost performance
P5	0.00	**1.39**	0.00	0.81	0.00	0.00	0.00	0.00
P6	0.00	0.93	0.00	0.93	1.03	0.00	0.00	0.00
P7	0.00	0.22	0.00	0.00	0.00	0.00	0.00	0.00
P8	0.00	0.85	0.00	0.00	0.00	0.00	0.00	0.00
P9	**1.12**	0.00	0.85	0.00	0.00	0.00	0.00	0.00
P10	1.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P11	0.84	0.00	0.00	0.00	0.00	0.00	1.00	0.00
P12	0.99	0.85	0.00	0.00	0.00	0.00	0.00	0.00
P13	0.00	0.00	0.73	0.00	0.00	0.00	0.00	0.00
P14	0.00	0.63	0.92	0.00	0.00	0.00	0.00	0.00
P15	0.00	0.00	0.00	0.00	0.69	0.00	**1.10**	0.00
P16	0.96	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P17	0.00	1.29	0.87	0.00	0.00	0.00	0.00	**0.86**
P18	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P20	0.00	0.00	0.00	0.00	0.85	0.00	0.00	0.00
P21	0.00	0.00	0.00	0.88	0.00	0.00	0.00	0.00
P22	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P23	0.00	0.00	1.07	0.00	0.64	0.00	0.00	0.00
P24	0.96	0.00	0.84	**1.02**	0.00	0.00	0.00	0.00

Appendix E1: PFSM of laptop dataset

	Screen	Customer service	Running speed	Startup speed	Appearance	Office work performance	Heat dissipation	Cost performance	Logistics	Weight	Keyboard	Configuration	game performance
w	0.0527	0.0358	0.2835	0.1063	0.0217	0.0425	0.0193	0.0640	0.0047	0.0072	0.0096	0.0241	0.0432
C1	0.26	0.04	0.14	0.59	0.26	0.00	0.83	0.00	0.00	0.00	0.00	0.00	0.00
C2	0.26	0.12	0.19	0.31	0.49	0.46	0.55	0.90	0.00	0.00	0.00	0.00	0.00
C3	0.49	0.68	0.24	0.59	0.57	0.00	**1.67**	**1.54**	**1.48**	0.90	0.00	0.00	0.00
C4	0.44	0.29	0.69	0.53	0.23	0.00	0.29	0.00	0.00	0.00	0.37	0.00	0.62
C5	0.24	0.21	0.21	0.14	0.34	0.46	0.00	0.00	1.11	0.00	0.00	0.00	0.00
C6	0.35	0.14	0.08	0.66	0.42	0.00	0.71	0.00	0.00	0.00	0.00	0.21	0.20
C7	0.40	0.59	0.41	0.43	0.61	0.00	0.00	0.46	0.97	0.00	0.53	0.55	0.36
C8	0.62	0.11	0.16	0.31	0.57	0.00	1.23	1.07	0.00	1.08	0.00	0.00	0.00
C9	0.53	0.60	0.17	0.63	0.32	**1.38**	0.00	1.34	0.00	1.34	**1.59**	0.00	0.00
C10	0.64	0.01	0.30	0.28	0.38	1.05	0.81	0.00	0.60	1.27	0.00	0.00	0.00
C11	0.27	**0.69**	0.38	0.47	0.52	0.17	0.00	0.00	0.00	0.83	0.00	0.00	0.00
C12	0.29	0.39	**0.75**	0.73	**0.68**	0.96	0.67	0.00	0.89	**1.46**	1.14	0.00	0.00
C13	0.05	0.15	0.30	0.37	0.46	1.10	0.23	0.00	0.00	0.00	1.04	0.98	0.16
C14	0.63	0.32	0.30	0.12	0.62	0.32	0.00	0.00	0.68	0.00	0.00	0.00	0.00
C15	**0.81**	0.65	0.37	**0.87**	0.57	0.69	0.00	0.00	0.00	0.96	1.42	0.00	0.00
C16	0.33	**− 0.05**	0.22	0.28	0.41	0.54	0.71	0.00	0.00	1.24	0.00	1.26	0.00
C17	0.21	0.23	0.42	0.52	0.50	0.00	0.90	1.34	0.00	0.00	0.00	1.17	0.00
C18	0.24	0.06	0.23	0.46	0.41	0.00	0.00	0.00	0.00	0.00	0.48	1.19	0.64
C19	0.39	0.53	0.38	0.34	0.60	0.43	0.78	1.09	0.47	0.00	0.00	**1.48**	0.00
C20	0.64	0.22	0.18	0.44	0.55	0.55	0.00	1.09	1.03	0.00	0.00	0.00	**0.68**

Appendix E2: PFSM of laptop dataset

	Price	Workmanship	System	Complimentary products	After-sales	Accessories	Quality	Office software	H-Share	Packaging	Easy to use	Audio
w	0.0503	0.0111	0.0228	0.0079	0.0206	0.0121	0.1007	0.0358	0.0095	0.0030	0.0076	0.0039
C1	0.38	0.00	0.00	**− 0.07**	**− 0.10**	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C2	0.00	0.00	0.00	0.00	0.00	**0.14**	0.00	0.00	0.00	0.00	0.00	0.00
C3	0.00	0.00	0.00	0.00	0.00	0.00	**1.18**	0.00	0.00	0.00	0.00	0.00
C4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.52**
C6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C9	0.00	0.00	**1.36**	0.00	**0.92**	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C10	0.00	0.00	0.38	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.83**	0.00	0.52	0.00
C12	0.00	0.77	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C14	0.31	0.63	0.00	**0.54**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C15	0.00	**1.30**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.81**	0.00
C16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.30**	0.00	0.00	0.00	0.00
C17	**0.48**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.83**	0.00	0.00
C19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Funding This work was supported by the National Natural Science Foundation of China (Grant Nos. 71872123, 72032005 and 71572122).

Declarations

Conflict of interest No potential conflict of interest was reported by the authors.

References

1. Gao, S., Tang, O., Wang, H., & Yin, P. (2018). Identifying competitors through comparative relation mining of online reviews in the restaurant industry. *International Journal of Hospitality Management, 71*, 19–32. https://doi.org/10.1016/j.ijhm.2017.09.004
2. Nasr, S. B., Bécan, G., Acher, M., Ferreira Filho, J. B., Sannier, N., Baudry, B., & Davril, J.-M. (2017). Automated extraction of product comparison matrices from informal product descriptions. *Journal of Systems and Software, 124*, 82–103. https://doi.org/10.1016/j.jss.2016.11.018
3. Liu, Y., Bi, J.-W., & Fan, Z.-P. (2017). Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. *Information Fusion, 36*, 149–161. https://doi.org/10.1016/j.inffus.2016.11.012
4. Amarouche, K., Benbrahim, H., & Kassou, I. (2015). Product opinion mining for competitive intelligence. *Procedia Computer Science, 73*, 358–365. https://doi.org/10.1016/j.procs.2015.12.004
5. Kim, S. G., & Kang, J. (2018). Analyzing the discriminative attributes of products using text mining focused on cosmetic reviews. *Information Processing & Management, 54*(6), 938–957. https://doi.org/10.1016/j.ipm.2018.06.003
6. Farhadloo, M., Patterson, R. A., & Rolland, E. (2016). Modeling customer satisfaction from unstructured data using a Bayesian approach. *Decision Support Systems, 90*, 1–11. https://doi.org/10.1016/j.dss.2016.06.010
7. Li, S., Zha, Z.-J., Ming, Z., Wang, M., Chua, T.-S., Guo, J., & Xu, W. (2011). Product comparison using Comparative Relations. Presented at the SIGIR '11 *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, Beijing, China*.
8. Zhang, Z., Guo, C., & Goes, P. (2013). Product comparison networks for competitive analysis of online word-of-mouth. *ACM Transactions on Management Information Systems, 3*(4), 1–22. https://doi.org/10.1145/2407740.2407744
9. Yang, X., Yang, G., & Wu, J. (2016). Integrating rich and heterogeneous information to design a ranking system for multiple products. *Decision Support Systems, 84*, 117–133. https://doi.org/10.1016/j.dss.2016.02.009
10. Tkachenko, M., & Lauw, H. W. (2017). Comparative relation generative model. *IEEE Transactions on Knowledge and Data Engineering, 29*(4), 771–783. https://doi.org/10.1109/tkde.2016.2640281
11. Kumar, A., & Abirami, S. (2018). Aspect-based opinion ranking framework for product reviews using a Spearman's rank correlation coefficient method. *Information Sciences, 460–461*, 23–41. https://doi.org/10.1016/j.ins.2018.05.003
12. Wang, Z., Zhao, Y., Xi, J., & Du, C. (2016). Fast ranking influential nodes in complex networks using a k-shell iteration factor. *Physica A: Statistical Mechanics and its Applications, 461*, 171–181. https://doi.org/10.1016/j.physa.2016.05.048
13. Wang, W. M., Wang, J. W., Li, Z., Tian, Z. G., & Tsui, E. (2019). Multiple affective attribute classification of online customer product reviews: A heuristic deep learning method for supporting Kansei engineering. *Engineering Applications of Artificial Intelligence, 85*, 33–45. https://doi.org/10.1016/j.engappai.2019.05.015
14. Zhang, L., Chu, X., & Xue, D. (2018). Identification of the to-be-improved product features based on online reviews for product redesign. *International Journal of Production Research, 57*(8), 2464–2479. https://doi.org/10.1080/00207543.2018.1521019
15. Wang, W. M., Li, Z., Tian, Z. G., Wang, J. W., & Cheng, M. N. (2018). Extracting and summarizing affective features and responses from online product descriptions and reviews: A Kansei text mining approach. *Engineering Applications of Artificial Intelligence, 73*, 149–162. https://doi.org/10.1016/j.engappai.2018.05.005

16. Candi, M., Jae, H., Makarem, S., & Mohan, M. (2017). Consumer responses to functional, aesthetic and symbolic product design in online reviews. *Journal of Business Research, 81*, 31–39. https://doi.org/10.1016/j.jbusres.2017.08.006
17. Jin, J., Ji, P., Liu, Y., & Johnson Lim, S. C. (2015). Translating online customer opinions into engineering characteristics in QFD: A probabilistic language analysis approach. *Engineering Applications of Artificial Intelligence, 41*, 115–127. https://doi.org/10.1016/j.engappai.2015.02.006
18. Abrahams, A. S., Jiao, J., Wang, G. A., & Fan, W. (2012). Vehicle defect discovery from social media. *Decision Support Systems, 54*(1), 87–97. https://doi.org/10.1016/j.dss.2012.04.005
19. Zhang, X., Niu, S., Zhang, D., Wang, G. A., & Fan, W. (2015). Predicting Vehicle Recalls with User-Generated Contents: A Text Mining Approach. *Presented at the Intelligence and Security Informatics: Pacific Asia Workshop, PAISI 2015, Ho Chi Minh City, Vietnam*.
20. Law, D., Gruss, R., & Abrahams, A. S. (2017). Automated defect discovery for dishwasher appliances from online consumer reviews. *Expert Systems with Applications, 67*, 84–94. https://doi.org/10.1016/j.eswa.2016.08.069
21. Mummalaneni, V., Gruss, R., Goldberg, D. M., Ehsani, J. P., & Abrahams, A. S. (2018). Social media analytics for quality surveillance and safety hazard detection in baby cribs. *Safety Science, 104*, 260–268. https://doi.org/10.1016/j.ssci.2018.01.003
22. Khare, V. R., & Chougule, R. (2012). Decision support for improved service effectiveness using domain aware text mining. *Knowledge-Based Systems, 33*, 29–40. https://doi.org/10.1016/j.knosys.2012.03.005
23. Korfiatis, N., Stamolampros, P., Kourouthanassis, P., & Sagiadinos, V. (2019). Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews. *Expert Systems with Applications, 116*, 472–486. https://doi.org/10.1016/j.eswa.2018.09.037
24. Su, Y., & Teng, W. (2018). Contemplating museums' service failure: Extracting the service quality dimensions of museums from negative on-line reviews. *Tourism Management, 69*, 214–222. https://doi.org/10.1016/j.tourman.2018.06.020
25. Du, J., Xu, H., & Huang, X. (2014). Box office prediction based on microblog. *Expert Systems with Applications, 41*(4), 1680–1689. https://doi.org/10.1016/j.eswa.2013.08.065
26. Hur, M., Kang, P., & Cho, S. (2016). Box-office forecasting based on sentiments of movie reviews and Independent subspace method. *Information Sciences, 372*, 608–624. https://doi.org/10.1016/j.ins.2016.08.027
27. Hua Li, Sourav S Bhowmick, & Aixin Sun. (2010). Affinity-Driven Prediction and Ranking of Products in Online Product Review Sites. *Presented at the Proceedings of the 19th ACM Conference on Information and Knowledge Management, Toronto, Ontario, Canada*.
28. Wang, H., & Wang, W. (2014). Product weakness finder: An opinion-aware system through sentiment analysis. *Industrial Management & Data Systems, 114*(8), 1301–1320. https://doi.org/10.1108/imds-05-2014-0159
29. Wang, W., Feng, Y., & Dai, W. (2018). Topic analysis of online reviews for two competitive products using latent Dirichlet allocation. *Electronic Commerce Research and Applications, 29*, 142–156. https://doi.org/10.1016/j.elerap.2018.04.003
30. Jin, J., Ji, P., & Gu, R. (2016). Identifying comparative customer requirements from product online reviews for competitor analysis. *Engineering Applications of Artificial Intelligence, 49*, 61–73. https://doi.org/10.1016/j.engappai.2015.12.005
31. Wasserman, S., & Faust, K. (1994). *Social network analysis: Mthods and applications*. Cambridge University Press.
32. Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., & Makse, H. A. (2010). Identification of influential spreaders in complex networks. *Nature Physics, 6*(11), 888–893. https://doi.org/10.1038/nphys1746
33. Garas, A., Schweitzer, F., & Havlin, S. (2012). A k-shell decomposition method for weighted networks. *New Journal of Physics, 14*(8), 083030. https://doi.org/10.1088/1367-2630/14/8/083030
34. Wei, B., Liu, J., Wei, D., Gao, C., & Deng, Y. (2015). Weighted k-shell decomposition for complex networks based on potential edge weights. *Physica A: Statistical Mechanics and its Applications, 420*, 277–283. https://doi.org/10.1016/j.physa.2014.11.012
35. Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science, 31*(3), 521–543. https://doi.org/10.1287/mksc.1120.0713

36. Chen, K., Luo, P., & Wang, H. (2016). Investigating transitive influences on WOM: From the product network perspective. *Electronic Commerce Research, 17*(1), 149–167. https://doi.org/10.1007/s10660-016-9241-8
37. Tan, K. H., Ji, G., Lim, C. P., & Tseng, M.-L. (2017). Using big data to make better decisions in the digital economy. *International Journal of Production Research, 55*(17), 4998–5000. https://doi.org/10.1080/00207543.2017.1331051
38. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research, 3*, 993–1022.
39. Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management, 59*, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009
40. Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management, 58*, 51–65. https://doi.org/10.1016/j.tourman.2016.10.001
41. Wang, Y., & Xu, W. (2018). Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. *Decision Support Systems, 105*, 87–95. https://doi.org/10.1016/j.dss.2017.11.001
42. Duan, W., Cao, Q., Yu, Y., & Levy, S. (2013). Mining Online User-Generated Content: Using Sentiment Analysis Technique to Study Hotel Service Quality. *Presented at the 46th Hawaii International Conference on System Sciences, Wailea, Maui, HI, USA*.
43. Lee, K., & Yu, C. (2018). Assessment of airport service quality: A complementary approach to measure perceived service quality based on Google reviews. *Journal of Air Transport Management, 71*, 28–44. https://doi.org/10.1016/j.jairtraman.2018.05.004
44. Bi, J.-W., Liu, Y., Fan, Z.-P., & Cambria, E. (2019). Modelling customer satisfaction from online reviews using ensemble neural network and effect-based Kano model. *International Journal of Production Research, 57*(22), 7068–7088. https://doi.org/10.1080/00207543.2019.1574989
45. Mirtalaie, M. A., Hussain, O. K., Chang, E., & Hussain, F. K. (2018). Extracting sentiment knowledge from pros/cons product reviews: Discovering features along with the polarity strength of their associated opinions. *Expert Systems with Applications, 114*, 267–288. https://doi.org/10.1016/j.eswa.2018.07.046
46. Eliacik, A. B., & Erdogan, N. (2018). Influential user weighted sentiment analysis on topic based microblogging community. *Expert Systems with Applications, 92*, 403–418. https://doi.org/10.1016/j.eswa.2017.10.006
47. Jeong, B., Yoon, J., & Lee, J.-M. (2017). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International Journal of Information Management*. https://doi.org/10.1016/j.ijinfomgt.2017.09.009
48. Jo, Y., & Oh, A. (2011). Aspect and sentiment unification model for online review analysis. *Presented at the WSDM '11 Proceedings of the fourth ACM international conference on Web search and data mining, HongKong, China*.
49. Qiu, J., Liu, C., Li, Y., & Lin, Z. (2018). Leveraging sentiment analysis at the aspects level to predict ratings of reviews. *Information Sciences, 451–452*, 295–309. https://doi.org/10.1016/j.ins.2018.04.009
50. Liu, B. (2015). *Sentiment analysis: Mining opinions*. Cambridge University Press.
51. Le, Q. V., & Mikolov, T. (2014). Distributed Representations of Sentences and Documents. *arXiv e-prints*. Retrieved from http://arxiv.org/abs/1405.4053
52. Elrod, T. (1991). Internal analysis of market structure: Recent developments and future prospects. *Marketing Letters, 2*(3), 253–266. https://doi.org/10.1007/BF00554130
53. Lee, T. Y., & BradLow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research (JMR), 48*(5), 881–894. https://doi.org/10.1509/jmkr.48.5.881
54. Saaty, T. L. (2001). Analytic hierarchy process. In S. I. Gass & C. M. Harris (Eds.), *Encyclopedia of Operations Research and Management Science* (pp. 19–28). New York, NY: Springer US. https://doi.org/10.1007/1-4020-0611-X_31
55. Saaty, T. L. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research, 48*(1), 9–26. https://doi.org/10.1016/0377-2217(90)90057-I