

Understanding Competition using Big Consumer Search Data

Daniel M. Ringel
Goethe University Frankfurt
dringel@wiwi.uni-frankfurt.de

Bernd Skiera
Goethe University Frankfurt
skiera@wiwi.uni-frankfurt.de

Abstract

Understanding competition is essential for every firm. Unfortunately, information on who competes with whom is scarce. Traditional research approaches such as surveys and panels are time consuming, expensive, and usually not well applicable to markets with large numbers of consumer durables. Consumer search data from a price comparison site, however, allows for observing consumers search and compare products, revealing which products they consider to be "competing" alternatives. We propose and critically compare three methods that use these data to identify, analyze and visualize competitive relationships between products. Finally, our empirical study of 147,336 consumers searching in the consumer electronics product category provides insight into consideration set size and composition, and shows that the average consideration set size of a consumer is 4.38 and the average consideration set of a product is 7.49.

1. Introduction

Understanding competition has long been recognized as essential for competitive strategy. It is the heart of strategic management decisions [4], supporting pricing policies, product design, product positioning and communication strategies [1],[21]. While sales of own products are directly observable, managers must resort to costly reports from market research firms for information on market shares and sales of competing products. The question of who their competitors are, however, remains unanswered.

To identify competitive relationships, research has turned to consumer perception and behavior. When faced with the decision of which product to purchase, consumers form a relevant set of alternative products, commonly referred to as consideration set [20]. As such, consideration sets provide time- and situation-specific information about consumer preferences and products they perceive as being competitors [18]. Since consumers determine which alternative products to include in their consideration sets, they become the ultimate arbiters of competitive relationships [17].

Past research has leveraged consideration sets to analyze competition and market structures. For example, DeSarbo et al. [4] identify competitive relationships and construct market structures based on survey data of jointly considered luxury cars as well as portable telephones. Similarly, Erdem [6] used panel data on repeat purchases of consumer products (e.g. yoghurt, detergent) to generate market maps.

Unfortunately, traditional approaches for competitive identification, such as surveys and consumer panel data, are usually prohibitively expensive and time consuming [11]. Moreover, they may suffer from sample biases, are confined to the cognitive boundaries of the interviewed consumers or require repeat purchases, which deem them inappropriate for consumer durable goods such as smartphones or TV-sets.

An attractive alternative are online search data of individual consumers. They can be used to identify consideration sets and thus competitive relationships between products (e.g. Moe [15]). Since online search data capture actual observed consumer behavior, they represent revealed measures of consumer search, which are more reliable than survey data [16]. Furthermore, online search data are not subject to the limitations of consumers' cognitive capabilities when trying to recall past consideration sets or comparing numerous alternatives in a test setting. And finally, technology allows for collecting information about these searches at rather low cost.

With millions of consumers searching and comparing millions of products online every day, big data on consumer online search activity should be available for nearly any consumer durable product. The challenge lies in capturing sufficiently representative online search data in a timely and cost effective manner, distilling the relevant consideration set information from it and generating valid and meaningful output that can easily be grasped by managers.

The aims of this article are threefold: Firstly, to propose three methods for identifying competing products, analyzing their relationships along metrics such as consumer interest (measured by search volume) and depicting the underlying market structure

in product consideration maps that incorporate product attributes and performance metrics. Secondly, to evaluate the suitability of the three methods for identifying, analyzing and visualizing competitive relations. Thirdly, to empirically investigate consideration set composition and size in consumer electronics and to compare the results of the three methods.

Our approach is unique as it is based on big consumer search data from a leading price comparison site that includes clickstream data from thousands of unique users in the LED-TV category. A major advantage of using price comparison site data is that it contains non-aggregated search activity of individual consumers and is not confined to the product catalog and customer base of a single retailer. Moreover, with over one million users searching for product information and prices every day, big consumer search data is available in real-time at low cost. And finally, big consumer search data from price comparison sites allow further analysis of consideration set size and composition as well as the development of new performance metrics for competitive analysis.

We organize the remainder of this article as follows: The next section provides an overview of relevant literature for consideration sets and their application in competitive analysis and market mapping. It is followed by a detailed description of the price comparison site data used in this study. We then propose three methods for analyzing competition. Subsequently, we empirically apply these methods to big consumer search data for the LED-TV category. Finally, we compare the results of the three methods and close with managerial implications arising out of our findings.

2. Related literature

The stream of past research devoted to competitor identification and market structure analysis lays testimony to the importance of understanding competitive relationships (e.g. Erdem [6], Bergen & Peteraf [1], and DeSarbo et al. [4]). Competitor identification and market structure analysis can be undertaken using different types of data such as consumer surveys, online product offerings [19] and panel-level scanner data [5] as well as different analytical methods such as cluster analysis (e.g. DeSarbo et al. [2]), multidimensional scaling (e.g. Erdem, [6]) and conjoint analysis (e.g. Green & Krieger [9]).

While some approaches to analyzing market structure and competitive relationships are based on

product attributes (e.g. Lee & Bradlow [12]), others draw on consideration sets of consumers (e.g. DeSarbo et al. [4]). We follow the latter approach since consumers become the ultimate arbiters of competitive relationships by selectively determining which alternative products to include in their consideration sets [17]. Consideration sets thus provide us with time- and situation-specific information about consumer preferences and products they perceive as competitors [18] and can thus be used to uncover underlying market structures [3]. However, unlike product attributes, consideration sets are latent constructs in the minds of consumers and are thus not directly observable [20].

Previous research has therefore resorted to either panel data (e.g. Erdem [6]) or survey data (e.g. DeSarbo [4]). Both data sources are however not ideal for the purpose of investigating consumer durable goods in markets consisting of many products. Panel data require repeat purchases common to fast moving consumer goods such as juice or detergents. However, consumer durables such as smartphones or TV-sets are purchased only once every few years. Surveys on the other hand are confined to the cognitive boundaries of consumers. This constraint limits surveys to a handful of alternative products that consumers are able to process at the same time and introduces uncertainty of whether consumers are able to correctly recall past purchase decisions or envision future purchase intent.

With the rapid spread and evolution of the World Wide Web, research has started to circumvent survey and panel data by leveraging user-generated content and online consumer search to analyze competitive relationships and market structures. For instance, Lee and Bradlow [12] use a text-mining algorithm to analyze online customer reviews for product attributes of digital cameras and apply clustering techniques to identify the underlying market structure. Kim et al. [11], on the other hand, use aggregated customer search data on camcorders from Amazon to estimate consideration sets as basis for mapping market structures using multidimensional scaling techniques. Both articles see a major advantage in using online data for market research since these data are less costly and timelier than surveys and panels.

In this article, we follow the notion of Kim et al. [11] to leverage online consumer search since it enables us to identify consideration sets and thus competitive relationships between products [15]. Moreover, online search data capture actual observed consumer behavior and therefore represent revealed measures of consumer search, which are regarded as more reliable than survey data [16].

Table 1. Overview of related research

| Features | Study | Erdem 1996 | DeSarbo et al. 2006 | Kim et al. 2011 | Lee & Bradlow 2011 | This Study |
|--|-------|------------|---------------------|-----------------|--------------------|------------|
| 1. Identification of Competitors | | ✓ | ✓ | ✓ | ✓ | ✓ |
| 2. Measurement of Competitor Strength | | | | | | ✓ |
| 3. Consideration of Product Attributes | | ✓ | | ✓ | ✓ | ✓ |
| 4. Visualization of Market Structure | | ✓ | ✓ | ✓ | ✓ | ✓ |
| 5. Applicable to Consumer Durables | | | ✓ | ✓ | ✓ | ✓ |
| 6. Use of Revealed Preferences | | ✓ | | ✓ | | ✓ |
| 7. Individual Consumer Data | | ✓ | ✓ | | ✓ | ✓ |
| 8. Number of Consumers | | 3,000 | 739 | n.a. | 8,526 | 147,336 |
| 9. Number of Products | | n.a. | 22 | 62 | 576 | 1,104 |

We are the first to use price comparison site data for observing, processing and analyzing the consideration set formation of individual consumers across a large number of products spanning an entire market. Further, we apply and compare multiple methods and propose performance metrics as basis for better understanding competitive relationships. Finally, we provide new insight into consideration set size and composition of competing consumer durables. Table 1 compares our study with previous research and outlines that our methods have more features and our empirical study covers more consumers and products than other studies.

3. Methodology

In this section, we provide an overview of the data, structure and function of price comparison sites. We then propose three methods, namely consideration ranking, consideration clustering and consideration mapping. Together, they allow us to identify, analyze and map competitive relationships.

3.1. Price comparison site data

Price comparison sites continue to gain popularity among consumers as they actively contribute to efficient markets and reduce search cost [7]. With millions of consumers searching and comparing hundred thousands of products in hundreds of product categories at price comparison sites, price comparison site data provide an encompassing representation of the market.

Being an intermediate, price comparison sites provide many value-adding functions such as search tools and an objective comparison of products and product offers [8] to a broad spectrum of consumers. Price comparison site data are thus by definition not confined to the inventory, customer base and sales policies (e.g. pricing and promotion) of a single retailer. Moreover, price comparison sites capture revealed measures of consumer search at a non-

aggregated, individual level, offering insight into individual customer journeys.

3.2. Structure of price comparison sites

To provide an overview of the general structure and content of price comparison sites and to introduce the price comparison site specific terminology used in this article, we provide a fictitious example of a typical user journey across a price comparison site.

In our example, Susan needs a new smartphone and decides to compare several models online. She starts by searching for a smartphone from Apple using Google. Google’s search results also include a link to a price comparison site, which Susan clicks on. She is taken to a product details page for the particular Apple smartphone at the price comparison site. We call this visit of a user to a product detail page a “page impression” for that particular product. Susan is presented with a general description of the smartphone, technical details, pictures, a product video, the average test score of several consumer tests, other user’s opinions and ratings of the product as well as a list of retail offers including price, availability, shipping cost, payment options, shop rating and a link to the online-shop of the respective retailer.

Susan studies the presented information and decides to take a look at an alternative smartphone from Samsung. This time she uses the search bar of the price comparison site to find the Samsung smartphone she is looking for. She is presented with a list of smartphones matching her search terms at the product listing page. Susan clicks on the Samsung smartphone she was looking for, opening its product details page (and generating a page impression for this Samsung smartphone model).

The following day, Susan continues her search and directly enters the URL of the price comparison site and navigates to the smartphone product category using the category navigation menu. To reduce the list of several hundred smartphones, Susan applies filters (e.g. display size, WiFi capability) and sorts the results by price. She then clicks on a model from HTC and is taken to its product details page (creating a page

impression for the HTC smartphone). Susan likes the HTC but wants to make sure it's the right choice for her.

Therefore, she revisits the product details pages of the Apple and Samsung smartphones (generating another page impression for each). She studies technical aspects of all products and eventually ends-up at the product details page of the HTC smartphone again (generating another page impression for it).

Susan then takes a closer look at the retail offers available and clicks on the offer with the second lowest price (the cheapest offer indicated that the smartphone was currently out of stock). A new browser tab/window opens with the website of the respective online-shop. We call this a "click" on a product offer.

For the analysis proposed in this article, we would summarize the relevant information of this customer journey as follows: Susan's consideration set size is 3. Each smartphone was viewed by one user (here Susan). Each smartphone received 2 page impressions. The HTC smartphone received 1 click, the others none. The conversion rate from page impressions to clicks (=CLK/PI) is 0.5 for the HTC smartphone and zero for the Apple and Samsung smartphones.

3.3. Data preparation

Before we can commence with our analysis we must process the clickstream data from the price comparison site. A clickstream consists of a series of actions a user (or consumer) undertakes along her customer journey across a website. We track these actions and log the sequence of products viewed (by page impression) along her journey. By allowing for reciprocity and transitivity among the products of a customer journey, we can construct consideration sets that include all products considered.

Table 2. Formation of consideration sets

| User | Clickstream | | Consideration Set | Set Size |
|---------|-------------|-----|-------------------|----------|
| Susan | D | E | D,E | 2 |
| Mike | B | A B | A,B | 2 |
| Jenny | C | B A | A,B,C | 3 |
| Tom | B | C B | B,C | 2 |
| Barbara | D | C | C,D | 2 |
| John | E | D | D,E | 2 |
| Lisa | A | C B | A,B,C | 3 |
| Chris | A | B | A,B | 2 |
| | | | Mean | 2.25 |

In a first step we translate individual user's clickstreams into a relational topology of jointly searched products normalized by the degree of relationship between them. To do so, we first use an automated procedure to process the log files containing the clickstreams of thousands of users

tracked at the price comparison site. The procedure identifies and analyzes each individual user's clickstream and translates product views (page impressions of product pages) into consideration sets (see Table 2).

As depicted in Table 3, the consideration sets are then organized in a joint consideration matrix (Y_{jk}) that expresses the relationship of joint consideration of product j and k observed in the clickstreams of all users.

Table 3. Resulting consideration matrix

| Joint product views | | | | | Summary statistics | | | | | | | |
|---------------------|---|---|---|---|--------------------|-------------------------------|-----|-----|-----|-----|-----|------|
| Y_{jk} | A | B | C | D | E | Product | A | B | C | D | E | Mean |
| A | 0 | 4 | 2 | 0 | 0 | Page impressions | 4 | 7 | 4 | 3 | 2 | 4.00 |
| B | 4 | 0 | 3 | 1 | 0 | Clicks on offers | 3 | 3 | 2 | 2 | 0 | 5.00 |
| C | 2 | 3 | 0 | 1 | 0 | Consideration set size (mean) | 2.5 | 2.4 | 2.5 | 2.0 | 2.0 | 2.28 |
| D | 0 | 0 | 1 | 0 | 2 | | | | | | | |
| E | 0 | 0 | 0 | 2 | 0 | | | | | | | |

Based on the notion that the degree of competitive relationship between two products is not only relative to how many times they are viewed together but also to the number of times they are each viewed, we apply Linden et al.'s [14] normalization procedure to normalize the degree of competitive relationship between two products j and k as follows:

$$(1) r_{jk} = \frac{n_{jk}}{\sqrt{n_j} \cdot \sqrt{n_k}}$$

where (r_{jk}) is the degree of competitive relationship, n_j (n_k) is the number of consumers who searched product j (k), and n_{jk} is the number of consumers who searched both j and k . Table 4 shows how the normalized relationships are mapped into matrix Y_{jk}^* .

Table 4. Normalized consideration matrix

| Normalized Joint Product Views | | | | | |
|--------------------------------|------|------|------|------|------|
| Y_{jk}^* | A | B | C | D | E |
| A | 0.00 | 0.76 | 0.50 | 0.00 | 0.00 |
| B | 0.76 | 0.00 | 0.57 | 0.22 | 0.00 |
| C | 0.50 | 0.57 | 0.00 | 0.29 | 0.00 |
| D | 0.00 | 0.00 | 0.29 | 0.00 | 0.82 |
| E | 0.00 | 0.00 | 0.00 | 0.82 | 0.00 |

3.4. Consideration set size

There are two perspectives on consideration set size that must be differentiated. The first, from a consumer point of view, expresses how many products a consumer jointly considers (see Table 2.). The second, from a product point of view, expresses how big the consideration sets are in which a given product is included (see Table 3.).

Table 5. Results of consideration ranking

| Product | Normalized Joint Consideration with Focal Product: A | Page Impressions | Clicks on Offers | Conversion Rate | Consideration Set Size |
|---------|---|------------------|------------------|-----------------|------------------------|
| A | 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 | 4 | 3 | 0.75 | 2.5 |
| B | 0.76 | 7 | 3 | 0.43 | 2.4 |
| C | 0.50 | 4 | 2 | 0.50 | 2.5 |
| D | 0.00 | 3 | 2 | 0.67 | 2.0 |
| E | 0.00 | 2 | 0 | 0.00 | 2.0 |
| Mean | | 4 | 2 | 0.47 | 2.28 |

In our numerical example the mean consideration set size of all consumers is 2.25 whereas for products it is 2.28.

3.5. Method 1: consideration ranking

We now conduct consideration ranking based on the normalized consideration matrix Y_{jk}^* (see Table 4) to identify the closest competitors of any given focal product. Since matrix Y_{jk}^* is in essence an inverse distance matrix of similarity (the greater the value, the more similar), we can directly identify and rank competitors of a focal product according to their normalized degree of joint consideration with the focal product. From Table 5 we see that product A’s nearest competitors are products B and C. As we observe no relationship to products D and E, we conclude that these products are not in competition with product A. We next analyze the competitive relationships in terms of how much interest each competing product attracts (number of page impressions) as well as with how many other products it is being considered with (consideration set size). Finally, we can also compare the conversion rate from page impressions to clicks on offers for each product.

As shown in Table 5, our competitive analysis indicates product B to be the closest competitor to product A. Furthermore, product B generates more consumer interest that is indicated by the larger number of page impressions. However, product B is not able to convert interest into clicks on offers as well as product A. Finally, product A typically faces slightly more competitors than product B as indicated by the slightly larger consideration set size.

3.6. Method 2: consideration clustering

While consideration ranking provides insight on competitors of one product, it offers little indication of which sets of products compete jointly together. We therefore employ hierarchical WARD-clustering to uncover joint competitive relationships of multiple

products. The clustering procedure is adapted to directly process our normalized joint consideration matrix Y_{jk}^* with similarity data as proximities for the identification of clusters.

The output of the hierarchical WARD-clustering is summarized in Table 6. The jump of the coefficient between stage 3 and 4 of the WARD linkage processing indicates that the optimal number of clusters is 2 (number of objects minus step number after which coefficient jumps = 5 – 3 = 2).

Table 6. Results of consideration clustering

| Stage | Cluster Combined | | Coefficient | Cluster Membership |
|-------|------------------|---------|-------------|--------------------|
| | CL 1 | CL 2 | | |
| 1 | 4 (D) | 5 (E) | 0.410 | (A) (B) (C) (D,E) |
| 2 | 1 (A) | 2 (B) | 0.790 | (A,B) (C) (D,E) |
| 3 | 1 (A,B) | 3 (C) | 1.020 | (A,B,C) (D,E) |
| 4 | 1 (A,B,C) | 4 (D,E) | 0.588 | (A,B,C,D,E) |

From the cluster membership column we see that in the case of two clusters, products A, B and C are competitors as well as products D and E.

3.7. Method 3: consideration mapping

Although, taken together, consideration ranking and consideration clustering paint a good picture of who competes with whom and how intense these competitive relationships are, they offer little insight into the underlying market structure and positioning a of all products relative to another. We therefore conduct consideration mapping to visualize the underlying market structure and transpose metrics of competitive performance as well as product attributes onto the generated consideration map.

In a first step we run multidimensional scaling procedures (PROXSCAL [13]) on our normalized joint consideration matrix Y_{jk}^* . The consideration map depicted in Figure 1 shows the close competitive relationship between products A and B as well as between products D and E as seen in our original normalized joint consideration matrix Y_{jk}^* (see Table

4.) Furthermore, product C is a more distant competitor to product A than B. This finding is also in line with the results of the output of the consideration clustering.

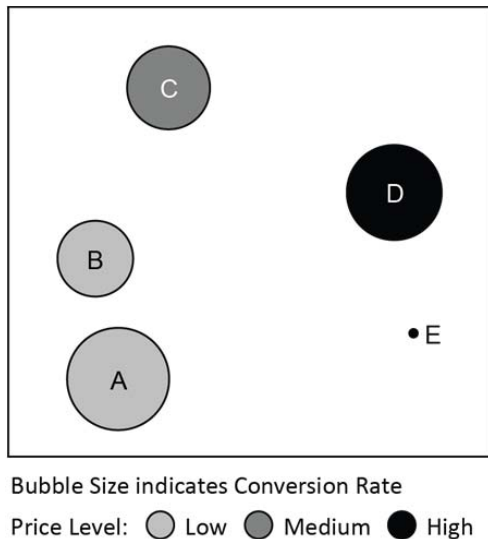


Figure 1. Results of consideration mapping

In the final step of our analysis, we explore competitive performance and drivers of market structure by transposing performance metrics and product attributes over the generated consideration map. For instance, we add conversion rates of the observed products to our consideration map by varying the bubble size of the products accordingly (the larger the bubble, the higher the conversion rate). Thus, our consideration map not only informs about relative positions of competitors but also about their strength (or performance) in terms of conversion rate (see Figure 1). Of course, other product attributes or metrics could also be used to represent bubble size and color.

4. Empirical study

In this section we apply the previously introduced methods to empirically demonstrate how they can process price comparison site data to better understand competition.

4.1. Data collection

We collected clickstream data in August 2012 from thousands of users in the LED-TV product category of a leading price comparison site. For the purpose of this study, a tracking pixel was developed and installed on all pages of the price comparison site. It allows us to

track each user along her journey and capture which products she looks at and on which offers she clicks. All user actions are captured with a timestamp that allows us to reconstruct the sequence of actions per user. A major advantage of obtaining consumer search data directly from the price comparison site is that we do not have to rely on error-prone crawler technology or published data summaries that may be indexed, normalized and/or truncated.

4.2. Data description

In our data we observe 147,336 users searching and comparing 1,104 products of 55 brands, generating 681,692 page impressions on products and 212,260 clicks on the offers of 269 online retailers. See Table 7 for further summary statistics.

Table 7. Summary statistics

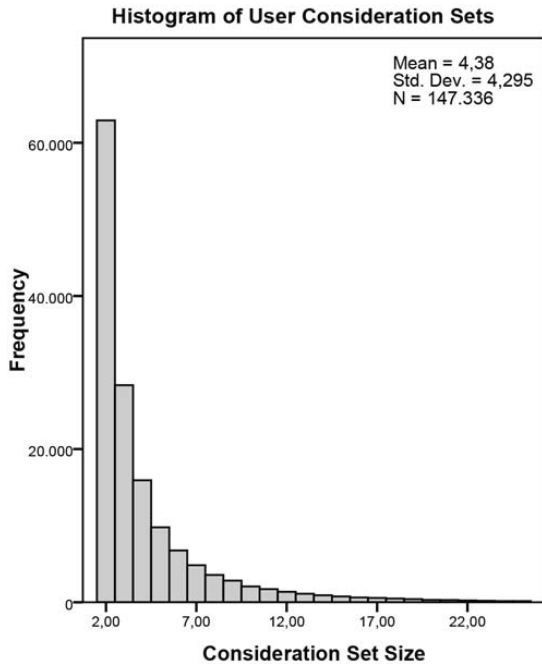
| | Page Impressions | Clicks on Offers |
|-----------|------------------|------------------|
| N | 1,104 | 1,052 |
| Min | 1 | 1 |
| Max | 18,241 | 5,235 |
| Sum | 681,692 | 212,260 |
| Mean | 617.47 | 201.77 |
| Std. Dev. | 1,425.89 | 442.55 |

4.3. Consideration set size

The average consideration set size (number of products considered) of users is 4.380 with a standard deviation of 4.29 and a positively skewed distribution (see Figure 2). This finding is in line with that of DeSarbo et al. [4], who, using survey data, found consideration sets of luxury cars to contain 2 to 6 models for the majority of consumers. On the other hand, Kim et al. [10] estimated the mean consideration set size for camcorders based on Amazon data to be 14. Both studies observe a long right tail, which we also observe in our data.

The average consideration set size a product is 7.490 with a standard deviation of 3.096 and a slightly positively skewed distribution (see Figure 2). Evidently there are some products that are considered with very few others, while at the other end there are products that are considered with many.

Overall we not only observe different product performance across metrics such as page impressions, and conversion rate, but also on consideration set size. We later use these metrics to compare products identified as competitors in order to get an indication of their relative competitive strength.



versus

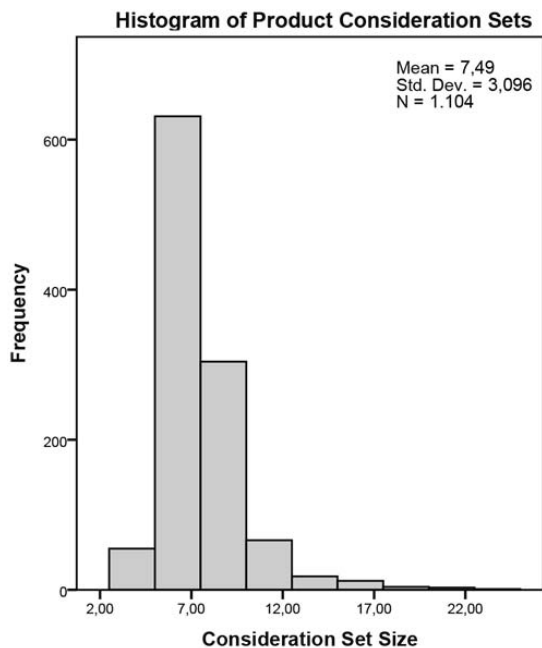


Figure 2. User vs. product consideration sets

4.4. Results of consideration ranking

Although the collected data span across 1,104 unique LED-TVs, it is not practical to demonstrate the proposed methods with such a large number of products. The output would be too encompassing to easily follow in this article (i.e. the joint consideration

matrix (Y_{jk}) would consist of over 1.2 million fields). Furthermore, the graphical consideration map would appear like a dense blot of undistinguishable dots.

We therefore demonstrate the three methods by randomly picking a focal product (Philips 40PFL5507K) and analyzing it with its top 50 most jointly viewed products. The resulting matrix Y_{jk} thus captures 1,250 relationships ($\frac{N \times N - N}{2}$) based on observed search behavior. We normalize these relationships as described in equation 1 into matrix Y_{jk}^* and proceed to applying our proposed methods.

Table 8 summarizes the results of the consideration ranking for the top 5 competitors of the focal product. The consideration ranking identifies the closest competitors to our focal product as being the Philips 40PFL5007K and Samsung UE40ES6300. Although the Samsung UE40ES6300 is jointly considered less often with our focal product than the Philips 40PFL5007K, it attracts more consumer interest (page impressions), has a slightly better conversion rate and a comparatively smaller consideration set size, making it the stronger competitor of the two. Furthermore, most of the closer competition to our focal product stems from same brand products (3 out of 5). While this finding suggests that brand may be a key differentiating attribute for consumers when considering which LED-TV to purchase, it also points to a potential inner-brand cannibalization problem of Philips.

4.5. Results of consideration clustering

We just summarize but do not display the details of the result of consideration clustering. From the agglomeration schedule of the WARD linkage processing we identify the number of optimal clusters to be 20. Cluster size ranges from 1 to 5 with a mean of 2.55. Again the output suggests that brand drives consideration set composition for LED-TVs since all but one cluster consists of a single brand. Our focal product is clustered with only one further competitor (Philips 40PFL5007K), which according to the consideration ranking, is also the competing product that is most often jointly considered.

4.6. Application of consideration mapping

We apply PROXSCAL multidimensional scaling procedures to the normalized joint consideration matrix Y_{jk}^* of the top 50 competitors of the focal product (Philips 40PFL5507K) to generate a two dimensional consideration map (see Figure 3.)

Table 8. Results of consideration ranking

| Product | Normalized Joint Consideration with Focal Product: Philips 40PFL5507K | Page Impressions | Clicks on Offers | Conversion Rate | Consideration Set Size |
|--------------------|--|------------------|------------------|-----------------|------------------------|
| Philips 40PFL5507K | | 10,435 | 3,299 | 0.32 | 4.64 |
| Philips 40PFL5007K | 0.589 | 4,211 | 1,101 | 0.26 | 4.42 |
| Samsung UE40ES6300 | 0.463 | 23,434 | 6,441 | 0.27 | 3.95 |
| Philips 40PFL5537K | 0.297 | 960 | 389 | 0.41 | 5.38 |
| Toshiba 40TL963G | 0.290 | 3,647 | 1,175 | 0.32 | 4.99 |
| Philips 40PFL6605H | 0.283 | 309 | 75 | 0.24 | 6.62 |
| Mean | | 7,166 | 2,080 | 0.30 | 5.00 |

In order to identify the closest competitors of the focal product easily from the consideration map, we draw a circle around the focal product encompassing exactly 5 products (including the focal product). We chose 5 products since the mean consideration set size of the focal product is 4.64 (see Table 8.). Both the consideration ranking and the consideration clustering indicate brand to be a key driver for joint consideration. We therefore transpose brand onto the consideration map and find that several formations (or groups) consist of single brands (as indicated by the ovals in Figure 4), confirming the findings of the previous analysis. In a final step, we transpose the performance metric page impressions (consumer interest) and the product attribute price (coded as low, medium and high) onto the plain consideration map to explore the underlying market structure (see Figure 4.).

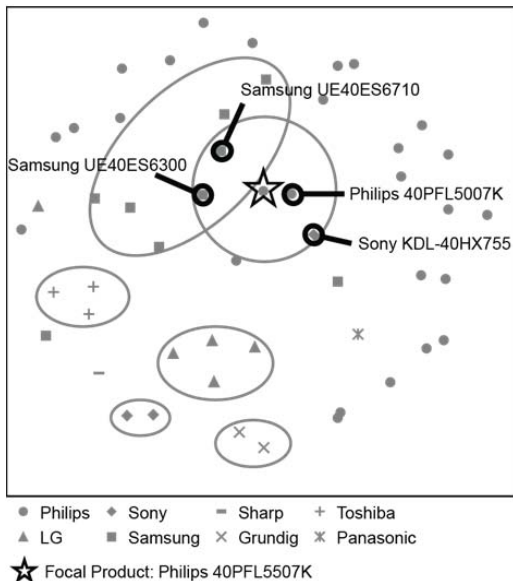


Figure 3. Results of consideration mapping

A close look at Figure 4 reveals that the focal product faces two close competitive products with a similar price (mid) but that both attract much more consumer interest (page impressions). There are also close competitors priced lower than the focal product as well as one closely competing product with a higher price attracting more consumer interest (page impressions). Finally, the consideration map indicates price to be a driver of market structure with higher priced products concentrated more towards the upper right and lower priced products concentrated more towards the bottom left.

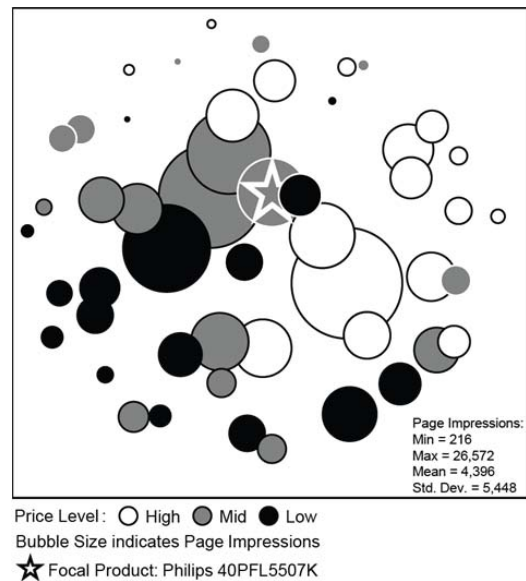


Figure 4. Results of consideration mapping

5. Comparison of methods

We now compare our proposed methods in regard to their suitability of answering the questions “Who am I competing with?” (identify competitors), “Which is my closest competitor?” (rank competitors), “Who

competes with whom in my market?” (identify groups of competitors)“ and “What does the market structure look like and what drives it?” (visualize market structure and its drivers).

Consideration ranking allows us to precisely identify and rank the closest competing products to our focal product since it is directly derived from the normalized joint consideration matrix. By adding metrics such as consumer interest (page impressions) to the consideration ranking we are able to also evaluate the strength of each competing product. However, a joint identification of competitors and a visualization of the market structure and its drivers cannot be achieved by consideration ranking.

Consideration clustering allows for a joint identification of competitors. However, it does not disclose which product is closest to a focal product and in our empirical study truncated the list of competitors of our focal product at one. Furthermore, consideration clustering does not visualize the underlying market structure and its drivers.

Consideration mapping produces a visual representation of all competing products relative to another. By transposing attributes onto the consideration map, it is possible to explore the drivers of the underlying market structure and of the consideration set composition of consumers. Although consideration maps provide a relatively good visual representation of which products compete most strongly, they produce no straight forward ranking or grouping of competing products.

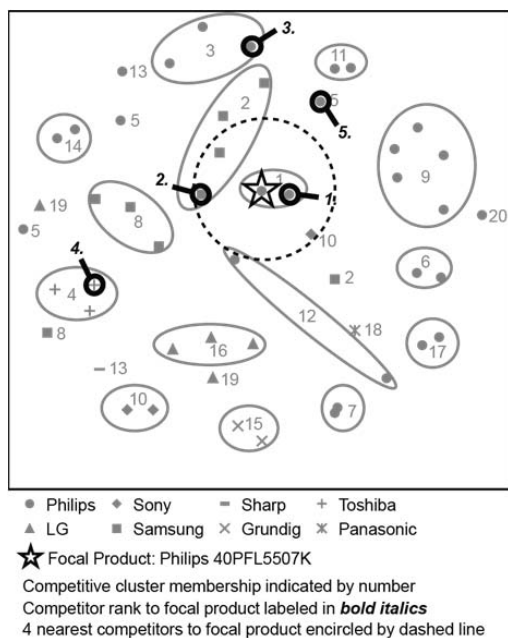


Figure 5. Method comparison

Figure 5 illustrates that all methods tend towards similar results. Most consideration clusters can also be identified as such in the consideration map and the closest competing product of the focal product is in the same cluster as well as nearest in the competitive map. Nevertheless, there are some discrepancies such as clusters 16 and 19, where the consideration map indicates that part of cluster 19 should in fact be in a competitive relationship with cluster 16. Also, the product identified through competitive ranking as being the 3rd closest competitor to the focal product is depicted as further away in the consideration map than for instance the product ranked 5th.

With each method having its own strengths and weaknesses (see Table 9.), only the combination of all three methods allows for an encompassing understanding the competitive environment.

Table 9. Comparison of methods

| | Consideration | | |
|--|---------------|------------|---------|
| | Ranking | Clustering | Mapping |
| Identify competing products | + | o | o |
| Rank competing products | + | - | o |
| Identify groups of competitors | - | + | o |
| Visualize market structure & drivers | - | - | + |
| Hit rate: % of top 4 competing products identified | 100% | 25% | 50% |

Suitability: + high o medium - not suitable

6. Conclusion

We set out to better understand competitive relationships by analyzing big consumer online search data from a price comparison site. We found that price comparison site data is an excellent source for this purpose. It provides revealed measures of consumer search from hundred thousands of consumers every day that spans across entire markets of consumer durable goods and provides several performance metrics (e.g. clicks, impressions) as well as product attribute data (e.g. brand, size). This characteristic makes them a highly valuable online data source that is timely and available at low cost.

Our study extends the existing research stream on competitive analysis and provides new insight into consideration set size and composition of competing consumer durables by using big online search data from a price comparison site as basis for observing,

processing and analyzing the consideration set formation of individual consumers across a large number of products spanning an entire market. Moreover, we propose and demonstrate the feasibility of three methods that all shed light on slightly different aspects of competition. The implications of using only a single method range from underestimating the potency of a competitive product to identifying the wrong products as being competitors. Managers are therefore well advised to employ a combination of the methods proposed in this article if they wish to attain a correct and encompassing understanding of the competitive situation faced by their brand and products.

We must however also recognize several limitations of our study. First, although all products considered by any given user are captured, repeat page impressions of a single product from the same user will be understated if the user opens a separate browser window for each product and switches between these windows instead of viewing them in sequence within a single window. Second, the formation of consideration sets may not only be driven by product attributes but also by factors that we are unable to observe in our data (e.g. advertising, word-of-mouth). Finally, recommendation features (e.g. "Top 10 also viewed products") at the price comparison site could artificially introduce products to a consumer's consideration set, causing a bias in our analysis. Yet, the results of Kim et al. [11] show that these effects are not too large.

7. References

- [1] Bergen, M., and Peteraf, M.A., "Competitor Identification and Competitor Analysis: A Broad-Based Managerial Approach", *Managerial and Decision Economics*, 23(4-5), 2002, pp. 157-169.
- [2] Desarbo, W., Howard, D., and Jedidi, K., "Multiclus: A New Method for Simultaneously Performing Multidimensional Scaling and Cluster Analysis", *Psychometrika*, 56(1), 1991, pp. 121-136.
- [3] Desarbo, W., and Jedidi, K., "The Spatial Representation of Heterogeneous Consideration Sets", *Marketing Science*, 14(3), 1995, pp. 326-342.
- [4] Desarbo, W.S., Grewal, R., and Wind, J., "Who Competes with Whom? A Demand-Based Perspective for Identifying and Representing Asymmetric Competition", *Strategic Management Journal*, 27(2), 2006, pp. 101-129.
- [5] Elrod, T., Russell, G., Shocker, A., Andrews, R., Bacon, L., Bayus, B., Carroll, J.D., Johnson, R., Kamakura, W., Lenk, P., Mazanec, J., Rao, V., and Shankar, V., "Inferring Market Structure from Customer Response to Competing and Complementary Products", *Marketing Letters*, 13(3), 2002, pp. 221-232.
- [6] Erdem, T., "A Dynamic Analysis of Market Structure Based on Panel Data", *Marketing Science*, 15(4), 1996, pp. 359-378.
- [7] Ghose, A., and Yao, Y., "Using Transaction Prices to Re-Examine Price Dispersion in Electronic Markets", *Information Systems Research*, 22(2), 2011, pp. 269-288.
- [8] Giaglis, G.M., Klein, S., and O'keefe, R.M., "The Role of Intermediaries in Electronic Marketplaces: Developing a Contingency Model", *Information Systems Journal*, 12(3), 2002, pp. 231-246.
- [9] Green, P.E., and Krieger, A.M., "Segmenting Markets with Conjoint Analysis", *Journal of Marketing*, 55(4)(4), 1991, pp. 20-31.
- [10] Kim, J.B., Albuquerque, P., and Bronnenberg, B.J., "Online Demand under Limited Consumer Search", *Marketing Science*, 29(6), 2010, pp. 1001-1023.
- [11] Kim, J.B., Albuquerque, P., and Bronnenberg, B.J., "Mapping Online Consumer Search", *Journal of Marketing Research*, 48(February), 2011, pp. 13-27.
- [12] Lee, T.Y., and Bradlow, E.T., "Automated Marketing Research Using Online Customer Reviews", *Journal of Marketing Research*, 48(5), 2011, pp. 881-894.
- [13] Leydesdorff, L., and Vaughan, L., "Co-Occurrence Matrices and Their Applications in Information Science: Extending Aca to the Web Environment", *Journal of the American Society for Information Science and Technology*, 57(12), 2006, pp. 1616-1628.
- [14] Linden, G.D., Smith, B.R., and Zada, N.K., "Use of Product Viewing Histories of Users to Identify Related Products", 06/28/2005 2005.
- [15] Moe, W.W., "An Empirical Two-Stage Choice Model with Varying Decision Rules Applied to Internet Clickstream Data", *Journal of Marketing Research*, 43(4), 2006, pp. 680-692.
- [16] Newman, J.W., and Lockeman, B.D., "Measuring Prepurchase Information Seeking", *Journal of Consumer Research*, 2(3), 1975, pp. 216-222.
- [17] Peter, J.P., and Olson, J.C., *Consumer Behavior and Marketing Strategy*, 3rd Ed., Irwin, Homewood, IL, 1993.
- [18] Roberts, J.H., and Lattin, J.M., "Development and Testing of a Model of Consideration Set Composition", *Journal of Marketing Research*, 28(4), 1991, pp. 429-440.
- [19] Sheng, Y.P., Peter P. Mykytyn, J., and Litecky, C.R., "Competitor Analysis and Its Defenses in the E-Marketplace", *Commun. ACM*, 48(8), 2005, pp. 107-112.
- [20] Shocker, A.D., Ben-Akiva, M., Boccara, B., and Nedungadi, P., "Consideration Set Influences on Consumer Decision-Making and Choice: Issues, Models, and Suggestions", *Marketing Letters*, 2(3), 1991, pp. 181-197.
- [21] Urban, G.L., and Hauser, J.R., *Design and Marketing of New Products*, Prentice-Hall, 2nd edn, Upper Saddle River, NJ, 1993.