Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Visualizing Asymmetric Competition Among More Than 1,000 Products Using Big Search Data

Daniel M. Ringel, Bernd Skiera

Marketing Science
Articles in Advance, pp. 1–24
ISSN 0732-2399 (print) | ISSN 1526-548X (online)
http://dx.doi.org/10.1287/mksc.2015.0950

Visualizing Asymmetric Competition Among More Than 1,000 Products Using Big Search Data

Daniel M. Ringel, Bernd Skiera
Department of Marketing, Faculty of Business and Economics, Goethe-Universität Frankfurt, 60629 Frankfurt am Main, Germany
{dringel@wiwi.uni-frankfurt.de, skiera@wiwi.uni-frankfurt.de}

In large markets comprising hundreds of products, comprehensive visualization of competitive market structures can be cumbersome and complex. Yet, as we show empirically, reduction of the analysis to smaller representative product sets can obscure important information. Herein we use big search data from a product- and price-comparison site to derive consideration sets of consumers that reflect competition between products. We integrate these data into a new modeling and two-dimensional mapping approach that enables the user to visualize asymmetric competition in large markets (>1,000 products) and to identify distinct submarkets. An empirical application to the LED-TV market, comprising 1,124 products and 56 brands, leads to valid and useful insights and shows that our method outperforms traditional models such as multidimensional scaling. Likewise, we demonstrate that big search data from product- and price-comparison sites provide higher external validity than search data from Google and Amazon.

Data, as supplemental material, are available at http://dx.doi.org/10.1287/mksc.2015.0950.

Keywords: big data; competitive market mapping; asymmetric competition; online search; product- and price-comparison sites

History: Received: December 30, 2013; accepted: May 28, 2015; Pradeep Chintagunta, Dominique Hanssens, and John Hauser served as the special issue editors and Peter Rossi served as associate editor for this article. Published online in *Articles in Advance*.

1. Introduction

Understanding competition and competitive market structure is essential for firms to derive a good competitive strategy (Rao and Sabavala 1986), which is at the heart of strategic management decisions (DeSarbo et al. 2006), supporting pricing policies, product design, product positioning, and communication strategies (DeSarbo et al. 1993, Urban et al. 1984, Bergen and Peteraf 2002, Lattin et al. 2003). Although managers can obtain some insight into the competitive landscape by analyzing their own product sales data or by purchasing reports on market shares, such information does not provide answers to the questions of who a firm's key competitors are and what the competitive structure of the market looks like. However, obtaining these answers is a complex undertaking, as the analyst must consider each competitor in relation to all others. Such a task is particularly difficult in large markets consisting of several dozen brands and hundreds of products.

Previous studies used information on brand switching (Erdem 1996), consideration sets (DeSarbo et al. 2006, DeSarbo and Grewal 2007), aggregated online search (Kim et al. 2011), and posts in discussion forums (Netzer et al. 2012) to uncover competitive relations among a limited number of products. These studies used perceptual mapping techniques to visualize these relations because such maps facilitate positioning decisions (Lilien and Rangaswamy 2004, Smelcer and Carmel 1997) and enhance decision quality (Ozimec et al. 2010).

A major challenge to the task of analyzing and visualizing competition is the growing number of competing products within product categories. For example, the category of LED television sets, which we focus on in what follows, comprises 1,124 products, corresponding to 56 brands. Traditional approaches for data collection, such as surveys and scanner panels, are not viable for product categories containing hundreds of products (Netzer et al. 2012). Surveys are limited by the cognitive capacity of interviewed consumers, who are unlikely to remember all products that they considered for purchase, whereas scanner panels require repeat purchases, making them inappropriate for consumer durables. Although previous research confined competitive analysis and visualization to a limited number of a priori selected products (e.g., DeSarbo and Grewal 2007, Kim et al. 2011), such confinement risks excluding relevant competing products from the resultant anterior definition of the market.

The aim of our manuscript is to develop, apply, compare, and empirically validate a new approach to analyze and visualize asymmetric competition in product categories comprising more than 1,000 products.

This new approach relies on big search data, specifically, clickstreams of thousands of consumers searching for and comparing products. It includes a newly developed model called *DRMABS* (*D*ecomposition and *R*eassembly of *MA*rkets *B*y *S*egmentation) that combines methods from multiple research disciplines such as biology, physics, computer science, and sociology with a new method of submarket-centric mapping to visualize asymmetric competition of more than 1,000 products in a single, two-dimensional map.

We base our approach on the notion that clickstream data of consumers searching for and comparing products online can be used to construct consideration sets (Moe 2006). Since consideration sets are the ultimate arbiters of competition (Peter and Olson 1993), they enable us to uncover competitive market structure (Roberts and Lattin 1991, DeSarbo and Jedidi 1995, Paulssen and Bagozzi 2006). An advantage of using individual clickstream data is that they capture actual observed behavior, which is more reliable than survey data (Newman and Lockeman 1975). Furthermore, online search data are not constrained by the cognitive capabilities of consumers, for example, by the extent to which consumers can recall past consideration sets. Finally, technology enables such data to be collected in real time at large volume and at low cost.

The remainder of this article is organized as follows: The next section provides an overview of related literature and previous approaches for competitive analysis and market mapping. Then we examine the need to analyze big data. We continue by outlining the basic idea of our approach and subsequently present a detailed description of how and why we combine and extend select methods in our model *DRMABS*. After an overview of the mechanics and the data available from product- and price-comparison sites, we externally validate search data and empirically apply our approach to the LED-TV market, including a comparison of *DRMABS* to existing models. We close with a discussion of key findings and implications for future research.

2. Previous Approaches

In this section we provide an overview of past studies aimed at analyzing and mapping competitive market structure and discuss their shortcomings.

2.1. Overview of Previous Approaches

As Table 1 shows, previous methods for mapping and analyzing competitive market structures have relied on data collected from scanner panels and surveys. For instance, Erdem (1996) uses scanner-panel data to build a dynamic market structure model to analyze competitive relations among brands of juice, detergent, and margarine. DeSarbo and Grewal (2007) use survey data on purchase intentions for cars as input for a multidimensional scaling model.

With the rapid spread and evolution of the Internet, researchers have turned to user-generated content and online consumer search in attempts to analyze competitive relationships and competitive market structure. Netzer et al. (2012) use text-mined user-generated content on automobiles from an online discussion forum to generate insights into the competitive landscape. Lee and Bradlow (2011) create a visualization of a competitive market structure on the basis of data on product attributes and brands' relative positions; these data are extracted automatically from text-mined online customer reviews. Kim et al. (2011) consider aggregated customer search data on camcorders from Amazon.com to derive a competitive market structure.

Table 1 also summarizes the methods that previous studies have used to visualize asymmetric competitive relationships. One approach focuses on the individual asymmetric competition between a focal product and each of its competing products; DeSarbo and Grewal (2007), for example, visualize these competitive relationships by duplicating each product in a common market map. Another approach is to visualize competitive asymmetry globally as a proxy for market share, across all products in a common market map by using different bubble sizes to indicate how often products are searched for by consumers (Kim et al. 2011). To avoid confusion, we refer to the asymmetric competitive relationship between two products as "local competitive asymmetry" and competitive asymmetry across all products as "global competitive asymmetry."

2.2. Shortcomings of Previous Approaches

Although previous approaches for visualizing competitive market structure have their merits, they also suffer from a number of shortcomings associated with data availability, visualization of large product categories, and the mapping of competitive asymmetry.

2.2.1. Shortcomings in Data Availability. Traditionally, data are collected either through purchase transactions, usually obtained from panels (e.g., Erdem 1996) or via surveys (e.g., DeSarbo and Grewal 2007). Neither data source is ideal for investigating consumer durables in large markets consisting of hundreds of products. Analysis of panel data requires repeat purchases, such as those common to fast-moving consumer goods (e.g., juice or detergents). However, a single consumer is not likely to purchase durables such as washing machines or TV sets more than once every few years. Survey data, in turn, are bounded by the cognitive capacity 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

Table 1 Overview of Previous Studies

Study	Erdem (1996)	DeSarbo and Grewal (2007)	Kim et al. (2011)	Lee and Bradlow (2011)	Netzer et al. (2012)	This study
Objective	Incorporate consumer choice dynamics into market structure models	Identify and represent asymmetric competitive market structures efficiently	Visualize online browsing behavior of consumers	Market structure visualization through automated interpretation of text-mined online customer reviews	Convert text-mined user-generated content into market structures and competitive landscape insights	Use of big search data to analyze and visualize asymmetric competition in categories containing over 1,000 durable products
Products	0	10	62	0	169	1,124
Brands	7	2	4	9	30	56
Consumers	3,000	289	Not applicable	Not applicable	76,587	105,606
Source	Scanner panel	Survey	Online retailer	Review platform	Discussion forum	Product- and price-comparison
Type of data	Brand switching	Consideration set	Summary of "also viewed products"	Product attribute association	Top of mind comention	Consideration set
Market definition	A priori	A priori	A priori	A posteriori	A posteriori	A posteriori
Visualization method	Dynamic market structure model	Asymmetric MDS	Asymmetric MDS	Correspondence analysis	Semantic network analysis (Kamada and Kawai), Classic MDS	*DRMABS*
Segmentation	—	—	—	*k*-means	Girvan–Newman, *k*-means	Multilevel Louvain
Normalization	Not applicable	Conditional probability	Salton cosine	Salton cosine	Lift	Mean conditional probability
Global asymmetry	—	—	Yes	—	—	Yes
Local asymmetry	—	Yes	—	—	—	Yes
Product attributes	Yes	—	Yes	Yes	Yes	Yes
Output	Common map	Common map with duplicate products	Common map	Common map	Common map	Common map
External validation	—	—	—	—	Yes ($r = 0.470$)	Yes ($r = 0.753$)
Real time	Yes	—	—	—	—	Yes
Low cost	—	—	Yes	Yes	Yes	Yes
Consumables	Yes	Yes	Yes	Yes	Yes	Yes
Durables	—	Yes	Yes	Yes	Yes	Yes
Large product categories	—	—	—	—	—	Yes

Notes. MDS, Multidimensional scaling; DRMABS, decomposition and reassembly of markets by segmentation.

future purchase intent. Moreover, surveys are costly and time consuming and do not reflect behavior in real time (Kim et al. 2011, Lee and Bradlow 2011, Netzer et al. 2012).

2.2.2. Shortcomings in Visualization. For a market containing a small number of products, it is possible to visualize the competitive market structure relatively easily by mapping dots onto a two-dimensional space, where each dot represents a single product. However, when the number of competing products increases, the graphical representation of these products quickly takes the form of a dense lump of dots, making the resulting map difficult to decipher (Netzer et al. 2012). Although a third dimension can be added to reduce this effect (e.g., DeSarbo and Grewal 2007), three-dimensional representations should be avoided wherever possible, since they complicate viewing and interpretation (Marbeau 1998).

Visual representations generated with multidimensional scaling techniques are particularly sensitive to the number of products being visualized: multidimensional scaling is inherently associated with a deterioration in the accuracy of product positions as the number of products increases (Buja et al. 2008, Faure and Natter 2010). Additionally, a circular bending effect, which refers to objects being mapped in a circular shape or "horseshoe," is common to multidimensional scaling solutions and can lead to an inaccurate interpretation of competitive relationships, since products that have weak or nonexistent competitive relationships with one another may appear closer together than they should (Kendall et al. 1970, Clark et al. 1986, Diaconis et al. 2008).

2.2.3. Shortcomings in Analysis of Competitive Asymmetry. Competitive asymmetry exists when the degree of competition between two firms is not equal, as when Firm A competes more intensely with Firm B than Firm B competes with Firm A (DeSarbo and Grewal 2007). For example, Apple is a large and well-known manufacturer of MP3 players (i.e., iPods), whereas iRiver only supplies a few models and is rather unknown. From iRiver's perspective, the competition with Apple is quite intense. From Apple's point of view, however, iRiver is hardly a competitor worth noting.

Past research has investigated asymmetric brand switching (e.g., Allenby and Rossi 1991), created indices of competitive asymmetry such as clout and vulnerability (e.g., Kamakura and Russell 1989), and discovered asymmetric structures in market data (e.g., Ramaswamy and DeSarbo 1990). Yet, the visualization of asymmetric relationships in competitive market structure maps has received relatively little attention.

DeSarbo et al. (2006) visualized asymmetric competition by using multiple competitive market structure maps, each reflecting the asymmetric relationship of a single focal product to its competing products in the market. The disadvantage of this approach is that the decision maker must consider a separate competitive market structure map for each product and evaluate it in comparison to all other maps, which becomes very cumbersome when the number of products is large. Recognizing this weakness, DeSarbo and Grewal (2007) developed a new model that depicts asymmetric relations by representing each product twice in a single map (once labeled in bold and once labeled in italics). The output map enables the observer to identify competitive asymmetry by looking at "italic-labeled" products relative to "bold-labeled" products. Although this approach eliminates the need to generate separate maps, its output becomes increasingly difficult to interpret as the number of products increases, since the visual representation becomes more and more cluttered with italic and bold labels.

Kim et al. (2011) visualized competitive asymmetry across all products in a single competitive market structure map by representing different products with bubbles whose size represents search frequency by online-shop users. The disadvantage of such techniques is that they enable the decision maker to observe only the overall competitive asymmetry across all products (i.e., global competitive asymmetry as a proxy for market share) and not the competitive asymmetry between any two individual products (i.e., local competitive asymmetry).

3. Analysis of the Need for Big Data

A major challenge to the task of analyzing and visualizing competition is the growing number of competing products within product categories. Past studies have limited their analysis to a small number of a priori selected products, such as 10 cars in the study of DeSarbo and Grewal (2007) or 62 camcorders in the study of Kim et al. (2011). Our approach, on the other hand, is specifically designed to handle hundreds to thousands of products that are linked to dozens of brands. We therefore aim to empirically investigate whether it is necessary to consider many products when analyzing and visualizing the competitive structure of large markets, or whether it is sufficient to account for only a few of the more important products.

For the purpose of this study, we acquired unit sales data from GfK that covers popular consumer durable markets in Germany, representing a wide range of product categories. We analyzed data for four product categories (washing machines, vacuum cleaners, digital cameras, and lawn mowers) to identify, for each category, the number of competing brands and products, market concentration, mean market share per product, the largest market share of a single product, cumulative market share of the top 50 products, number of brands included in the top 50 products, and the most products of a single brand in the market (see Table 2).

The consumer durable markets investigated consist, on average, of more than 1,000 products and 55 brands each. Using the Herfindahl–Hirschman index (HHI) to measure market concentration, we find that all markets are characterized by relatively low concentration, indicating a highly competitive environment. We also find that, in each market, the market share of the most dominant product is fairly low. More importantly, we find that if an analysis of the competitive market structure in these large markets were to be confined to the top 50 products (in terms of market share), then on average almost 50% of the sold products and 77% of the competing brands would not be accounted for.

We now take a closer look at the market composition of the LED-TV market, which we later use in the empirical application of our new approach (see §7). We find a total of 1,124 products from 56 brands in the LED-TV market in September 2012. The HHI of this market is only 50, indicating a highly competitive environment. By sorting products in ascending order according to market share, we find that an analysis confined to the top 50 LED-TVs would exclude a combined market share of over 60% for all of the remaining products and would eliminate 90% of all brands that are present in the LED-TV market (see Figure 1).

These findings suggest that, in the five markets discussed above (and in other large markets characterized by a similar structure), an analysis confined to 10 or even 50 a priori selected products would exclude large portions of the market. Moreover, such exclusion would make the findings of the analysis of the competitive market structure analysis useless for smaller manufacturers whose products are not among the top 50

Table 2 Characteristics of Four Consumer Durable Markets

Market[a]	Brands	Products	HHI[b]	Largest product share (%)	Mean product share (%)	Top 50 cumulative (%)	Brands in top 50	Most products of a brand
Washing machines	43	1,196	54	2.23	0.08	40.57	15 (35%)	151 (Siemens)
Vacuum cleaners	96	1,514	65	3.01	0.07	43.60	12 (13%)	117 (Miele)
Digital cameras	48	920	98	3.98	0.11	55.99	10 (21%)	102 (Fujifilm)
Lawn mowers	33	518	140	5.18	0.19	61.34	8 (24%)	69 (Wolf)
Mean	**55**	**1,037**	**89**	**3.60**	**0.11**	**50.38**	**11 (23%)**	**110**

Notes. HHI < 100 indicates a highly competitive market, and HHI $< 1{,}000$ indicates that the market is not concentrated. HHI, Herfindahl–Hirschman index.
[a]Source: GfK retail panel Germany (September 2012).
[b]Herfindahl–Hirschman index measures market concentration of products (1 to 10,000).

(e.g., manufacturers of newly launched brands, luxury brands, and niche brands). Managers of firms whose products are included among the top 50 products, on the other hand, might draw false conclusions about their competitive situation when the analysis is based solely on a handful of a priori selected products, since competitive threats from smaller, evolving manufacturers will remain undetected for a long time.

We therefore conclude that competitive market structure analysis of today's consumer durable markets requires an approach that is capable of collecting and analyzing at least 1,000 products and 55 brands, which provides a strong argument for the use of big data and approaches such as the one we propose next.

4. Description of our Approach

We now propose a new approach that enables asymmetric competition in large product categories—i.e., categories containing over 1,000 products—to be visualized in a single competitive-market-structure map. Our approach consists of the following five phases (see Figure 2): (i) collect data by observing consumers' behavior on a website that supplies information on many different products; (ii) consider products that are viewed together by the same consumer as products that are in that consumer's consideration set; (iii) identify competitive asymmetry; (iv) use our newly developed model *DRMABS* to derive a single, low-dimensional map that visualizes asymmetric competitive market structure; and (v) transpose product attributes onto the derived map.

4.1. Phase 1: Data Collection

We start in phase 1 with the collection of big search data (individual clickstreams) of thousands of consumers. Such search data can be obtained from any website that offers information on a broad range of products;

Figure 1 (Color online) Market Concentration of the LED-TV Market in September 2012

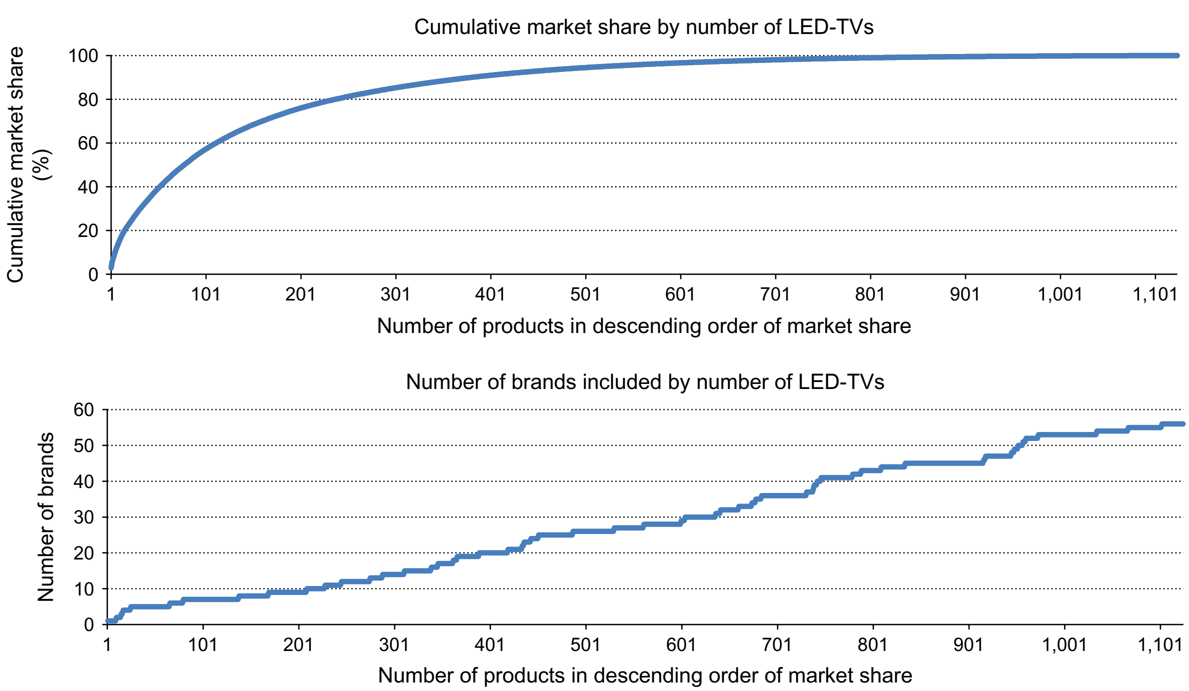
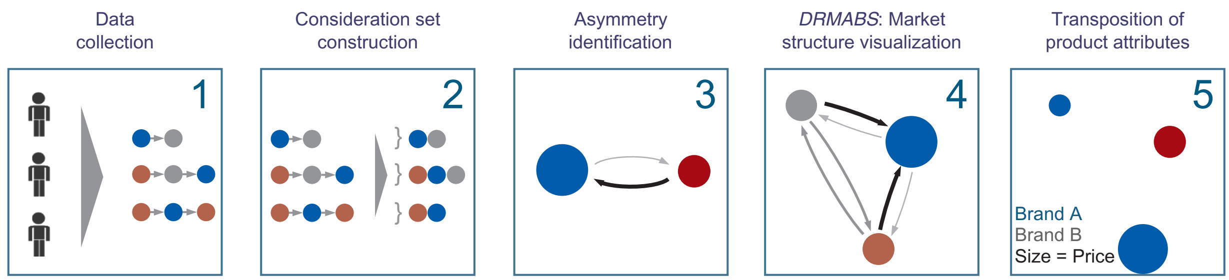


Figure 2 **Phases of Our Approach**



such websites include online retailers, product- and price-comparison sites, and product review sites.

Traditionally, cookies are used to capture user information such as clickstreams on websites. However, cookies recording clickstreams are prone to being rejected by browsers or deleted regularly. Tracking pixels constitute a better alternative for collecting clickstream data (Error and Error 2004). A tracking pixel is a tiny, usually transparent image that is embedded in the HTML code of a website to track usage by website visitors (Sipior et al. 2011). Every time a consumer opens a new webpage on the site, her browser also requests and downloads the tracking pixel from the Web server, allowing the Web server to notice and log the fact that the consumer has visited that specific page.

Moreover, additional information from the browser can be passed to the Web server when the request to load the tracking pixel is made, allowing the Web server to gather information on the conditions under which the request was made (e.g., which filters were set by the consumer) as well as track the consumer herself over time through an identifier.

In our approach, we capture the clickstreams of consumers searching for and comparing products on a website using a tracking pixel that we specifically developed for our empirical studies.

4.2. Phase 2: Construction of Consideration Sets

In phase 2 we construct consideration sets from the data collected in phase 1. An individual's consideration set comprises the products that she perceives as competitors (Roberts and Lattin 1991). Consideration sets can be constructed from individual choice data (e.g., Swait and Erdem 2007) as well as from clickstream data of consumers searching for and comparing products online (Moe 2006). Once constructed, consideration sets can be used to uncover competitive market structure (Urban et al. 1993, DeSarbo and Jedidi 1995), that is, the sets of products that customers consider to be substitutes for one another in specific usage situations.

We construct a consideration set for each consumer by analyzing her clickstream data with regard to the products she searched for and compared online (Moe 2006). The set of products that consumer i visits is denoted U_i; U_i is a subset of the universal set of products U of the corresponding product category ($U_i \subseteq U$). We aggregate the consideration sets across all consumers into a symmetric dyad matrix Y of joint product consideration. Specifically, in this matrix, Y_{jk} represents the number of consumers who considered products j and k together

$$Y_{jk} = \sum_{i \in I} G_{ij} \times G_{ik}, \quad (1)$$

where I is the index set of consumers, and the inclusion of products j and k in a consideration set U_i is operationalized as $G_{ij} = (j \in U_i \to 1, 0)$ and $G_{ik} = (k \in U_i \to 1, 0)$, respectively.

We subsequently use a numerical example (shown in Table 3) based on the clickstreams of eight consumers, who clicked on several products denoted A through E. These clickstreams are then used to construct an individual consideration set for each consumer. The consideration sets of all eight consumers are summarized in a symmetric dyad matrix Y of joint product consideration (see Table 3).

Table 3 suggests that a consideration set can be defined according to two different "perspectives": a consumer perspective and a product perspective. Specifically, there is a difference between the mean size of a consumer consideration set, i.e., the number of products being considered by the average consumer (2.25 products in our example), and the mean size of a product consideration set, i.e., the average size of the consideration sets in which the product is included (2.28 products).

4.3. Phase 3: Identification of Competitive Asymmetry

Our approach captures competitive asymmetry and distinguishes between competitive asymmetry across all products in a market (global competitive asymmetry) and direct competitive asymmetry between pairs of products (local competitive asymmetry).

4.3.1. Global Competitive Asymmetry. We define global competitive asymmetry as consumers' overall propensity to consider certain products, which serves as a proxy for market share. The basic idea of global

Table 3 Numerical Example

Consumer	Clickstream			Consideration set	Consumer consideration 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

Product	Consideration set inclusion (A_j)	Average product consideration set size
A	4	2.5
B	5	2.4
C	4	2.5
D	3	2.0
E	2	2.0
Mean	3.6	2.28

Joint consideration (Y)

Y_{jk}	A	B	C	D	E
A		4	2	0	0
B	4		3	0	0
C	2	3		1	0
D	0	0	1		2
E	0	0	0	2	

Global competitive asymmetry (A_j)
Frequency of consideration

A_j	
A	4
B	5
C	4
D	3
E	2

Local competitive asymmetry (Y^*)
Conditional probability of joint consideration

Y^*_{jk}	A	B	C	D	E
A		1.00	0.50	0.00	0.00
B	0.80		0.60	0.00	0.00
C	0.50	0.75		0.25	0.00
D	0.00	0.00	0.33		0.67
E	0.00	0.00	0.00	1.00	

competitive asymmetry is that if a given product A is considered by more consumers than products B, C, and D, then product A is the stronger competitor overall. We therefore not only identify how many products are competing but also determine how consumer consideration is distributed across these competing products. We capture global competitive asymmetry in vector $\mathbf{A}$, which contains, for each product j, the number of consideration sets in which j is included, as follows:

$$\mathbf{A}_j = \sum_{i\in I} G_{ij}, \tag{2}$$

where $G_{ij} = (j \in U_i \to 1, 0)$.

4.3.2. Local Competitive Asymmetry. As discussed in §2.2.3, local competitive asymmetry between pairs of products reflects the case in which one product is a more intense competitor to the other than vice versa. For instance, imagine that product A is included in the consideration sets of 40 consumers, and product B is included in the consideration sets of 10 consumers. A and B are jointly considered by five consumers. Therefore, product A is a competitor to product B in 50% (5/10) of its consideration sets, whereas product B is a competitor to product A in only 12.5% (5/40) of its consideration sets. Thus, product A is a more intense competitor to product B than vice versa. We capture local competitive asymmetry through conditional probability as follows:

$$P(k \mid j) = \frac{P(j \cap k)}{P(j)}, \tag{3}$$

where $P(k \mid j)$ is the probability of product k being considered given that product j is considered. We now define an asymmetric dyad matrix Y^*, in which Y^*_{jk} represents the conditional probability of joint consideration $P(k \mid j)$, across all consumers

$$Y^*_{jk} = \frac{\sum_{i\in I} G_{ij} \times G_{ik}}{\sum_{i\in I} G_{ij}}, \tag{4}$$

where $G_{ij} = (j \in U_i \to 1, 0)$ and $G_{ik} = (k \in U_i \to 1, 0)$.

Our numerical example shown in Table 3 provides a comparison of the associated symmetric matrix Y of (unconditional) joint consideration with the corresponding asymmetric matrix Y^* of conditional joint consideration.

By contrast to the symmetric matrix Y, the asymmetric matrix Y^* is directional, as indicated by the fact that the values of Y^*_{jk} and Y^*_{kj} are not all identical. By reading across row j, we find the conditional probability for each competing product k to be considered jointly with product j. For instance, for product C of matrix Y^* in Table 3, the most intense competitor is product B (0.75), followed by products A (0.50) and D (0.25).

4.4. Phase 4: Visualization of Asymmetric Competitive Market Structure

We visualize asymmetric competitive market structure in large product categories using a combination of adapted and extended methods from the areas of network analysis and graph theory. Network analysis and graphing methods have been used successfully to analyze and visualize biological networks (Girvan and Newman 2002), investigate the taxonomy of financial portfolios (Onnela et al. 2003), visualize human brain functional networks (Meunier et al. 2009), conduct medical research on vaccines (Rappuoli and Aderem 2011), and understand the behavior of large-scale systems for the purpose of designing and validating new Internet protocols and applications (Yao and Fahmy 2014).

In marketing, network measures such as degree and betweenness are frequently used to characterize the positions of individuals in social networks such as Facebook (e.g., Hinz et al. 2011), but a network approach like ours has not been used to visualize asymmetric competitive relations between products.

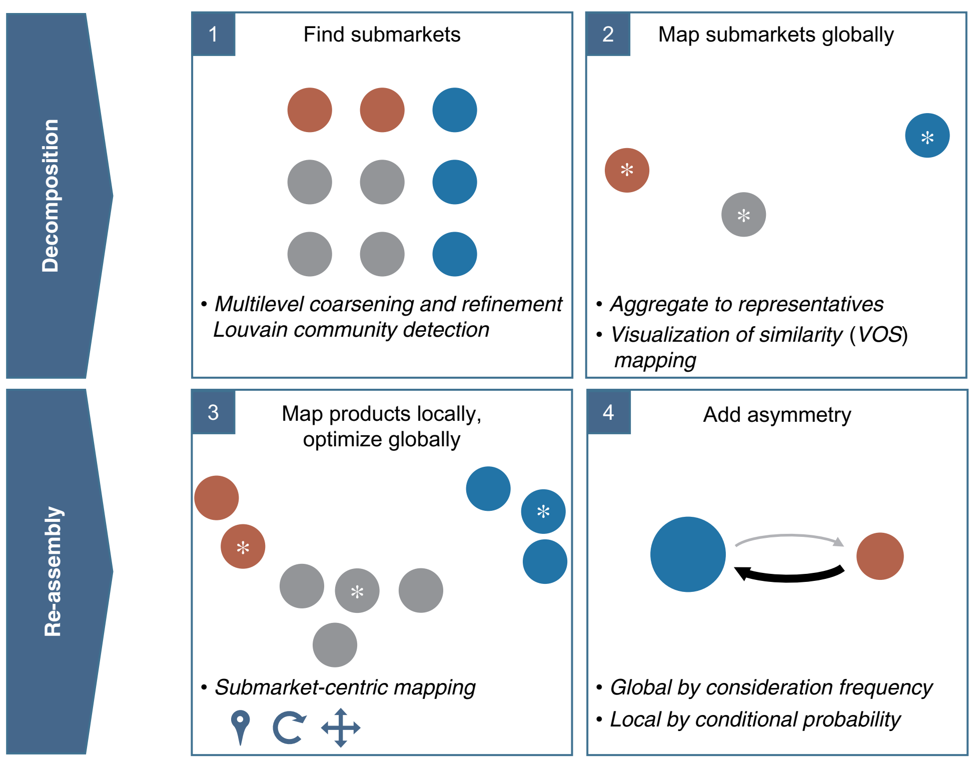
In network analysis, objects, in our case products, are referred to as vertices, which are connected by edges, which represent relationships among vertices. In our approach, an edge between two products represents the co-occurrence of these products in consumers' consideration sets. More frequent co-occurrence of two products—i.e., a stronger relationship—is indicated by greater edge weight. In a visual representation, vertices connected by heavier edges are positioned closer to each other compared with vertices connected by lighter edges. We consider edges to be undirected, reflecting symmetric relationships between vertices. To represent the asymmetric relations between each pair of products—in our case, the conditional probability that one product is considered in a set given that another is included in the set—we use two (directed) arcs with weights that correspond to the (asymmetric) conditional probabilities of the two products being analyzed.

Our objective is to combine network analysis and graphing methods to visualize the asymmetric competitive market structure for a product category comprising over 1,000 products. This visualization represents a challenge in markets with large numbers of products because groups of mapped objects tend to overlap in the resulting visual representation of the network, making relationships between individual products difficult to decipher (Palla et al. 2005). Furthermore, in visualizing the market, we seek to be able to determine whether all products in a market compete with one another or whether "submarkets" (clusters) exist with high levels of competition within them and low levels of competition among them. This goal is driven by the observation that correct partitioning of a market into submarkets explains more about consumer behavior than is apparent from the unpartitioned market (Urban et al. 1984).

To overcome the problem of overlapping objects and to facilitate identification of submarkets, we develop a new model called *DRMABS* (*D*ecomposition and *R*eassembly of *MA*rkets *B*y *S*egmentation), that combines and extends advanced segmentation and mapping methods. The underlying principle of *DRMABS* is that it decomposes the total market into submarkets and determines their relative locations before again reassembling local submarket structures into a global representation of the full market. Figure 3 provides an overview of the four main steps of *DRMABS*.

In step 1, we identify submarkets that, taken together, make up the entire competitive landscape of the market

Figure 3 **Four-Step Procedure of New Model *DRMABS* for Visualizing Asymmetric Competitive Market Structure**



under analysis. To do so, we generate a coarse-grained representation of the entire network of competing products that consists of product submarkets, where a submarket is defined as a group of products that compete intensely among themselves and weakly with products outside the group.

In step 2, we select one representative product for each identified submarket and aggregate all between-submarket relations (edges), such that each edge connecting two products in different submarkets is reassigned to the representatives of those submarkets. This step reduces the full network of all products to a smaller network of submarket representatives only. We then map these submarket representatives in relation to one another based on the aggregated edge weights to determine the relative positions of all identified submarkets.

In step 3, we conduct a new submarket-centric mapping that maps submarkets locally and also optimizes them globally across all products to obtain a nonoverlapping network configuration for the entire competitive market map. Our new submarket-centric mapping method preserves submarket locations relative to one another, rotates submarkets so that between-submarket relations are observed, and rescales each submarket to a common scale across all submarkets.

In step 4, we introduce global competitive asymmetry using bubble size to represent frequency of consideration as a proxy for marketing share, and local competitive asymmetry using weighted arrows to create a full asymmetric competitive market structure map.

We describe the methods and algorithms combined in our model *DRMABS* as well as the reasoning behind each choice in detail in §§4.4.1–4.4.4 and provide a summary overview in Table 4.

4.4.1. Step 1: Find Submarkets. We use co-occurrence data of all products to (i) identify the number of existing submarkets (if any); (ii) identify the products belonging to each submarket; and (iii) obtain an indication of the quality of our segmentation.

WARD-clustering and *k*-means partitioning methods are commonly used by marketers to segment markets. Although these methods are convenient and widely available in statistical software packages, they are not fully adequate for our purposes owing to several weaknesses. First, these methods require the number of submarkets to be provided either a priori (*k*-means) or a posteriori (e.g., elbow criteria for WARD-clustering). Second, they are restricted in their capacity to indicate the quality of their solutions. Third, these approaches are usually used to consider small numbers of products, whereas our goal is to consider more than 1,000 products. Finally, hierarchical methods such as WARD-clustering have difficulty in correctly detecting small submarkets (Fortunato and Barthelemy 2007), and *k*-means partitioning handles outlying products poorly (Newman 2004).

Table 4 Method Selection

Steps	Method selected	Reasoning	Previous studies	Origin
1. Find submarkets	Louvain community detection with resolution parameter	• Number of clusters determined a posteriori • Community quality optimization • No resolution limit	Newman (2004) Fortunato and Barthelemy (2007) Blondel et al. (2008)	Physics, Biology, Mathematics
	Multilevel coarsening and refinement clustering	• Outperforms standard modularity optimization • Not greedy: no prematurely merged clusters • Handles large networks with very heterogeneous cluster size	Lancichinetti and Fortunato (2009) Rotta and Noack (2011)	Physics, Computer Science
2. Map submarkets globally	Harmonic centrality	• Identifies the most centrally located object in cluster • Considers indirect relationships • Captures unconnected objects	Boldi and Vigna (2014)	Graph Theory, Mathematics
	Visualization of similarities	• No circular bending • No lumping of prominent objects • Specific for co-occurrence data • Captures indirect relations	Diaconis et al. (2008) van Eck and Waltman (2007)	Bibliometrics
3. Map products locally, optimize globally	Submarket-centric mapping	• Few overlaps • Maps submarkets individually within full network • Preserves relative positions of submarkets • Proportional scaling and optimal submarket rotation	Palla et al. (2005) De Nooy et al. (2011)	This study
4. Add asymmetry	Conditional probability consideration frequency	• Competition is asymmetric • Differentiate between global asymmetry across all competitive relationships vs. local asymmetry between pairs of competitors	DeSarbo et al. (2006) Kim et al. (2011)	Marketing

We therefore turn to another research discipline, namely, network analysis, to meet our segmentation aims. Netzer et al. (2012) recently demonstrated the use of a community detection method developed by Girvan and Newman (2002) to derive submarkets for a market consisting of 169 cars. A community is defined as a subgraph of a network whose vertices are more tightly connected to one another than to vertices outside the subgraph. When edges connecting vertices have weights, these weights are also taken into account in the formation of communities (Newman and Girvan 2004). The aim of community detection is therefore to identify distinct submarkets whose inner relationships are stronger than their outer relationships.

From a technical perspective, community detection is a greedy optimization of what is called the "modularity" of communities in a network (Fortunato and Barthelemy 2007). Modularity is a measure that compares the number of links inside a given subnetwork (module) with the expected value for a randomized graph of the same size and same degree sequence. Thus, modularity is a quantitative measure for the quality of the division of a network into communities (Newman 2004), or in our case, the division of a market into submarkets.

Beyond the advantage of offering an indication of quality, the community detection approach also determines the number of communities (here submarkets) without a priori input and is capable of directly processing large co-occurrence data. As such, community detection helps us to reach all of our segmentation aims. Yet, basic modularity optimization as proposed by Newman (2004) can miss important substructures of a network, which happens frequently when networks are very large and contain communities with different sizes (Fortunato and Barthelemy 2007).

To ensure that we do not miss any important substructures, we combine the following two methods into a single procedure: (i) a modified version of the Louvain community detection method proposed by Blondel et al. (2008) and (ii) a multilevel coarsening and refinement procedure developed by Rotta and Noack (2011). The advantage of our modification of the Louvain community detection method is that it contains a resolution parameter to influence the level of granularity of the solution, enabling us to arrive at a solution that fits the underlying structure of our data. The advantage of adding a multilevel coarsening and refinement procedure to community detection is that it counters the greedy nature of modularity optimization. That means that it can prevent the premature merging of submarkets and thus the failure to detect substructures correctly (Rotta and Noack 2011).

The resulting integrated combination of methods outperforms common methods for segmenting large networks, especially since it is able to detect comparatively small submarkets in relatively large networks (Lancichinetti and Fortunato 2009, Fortunato 2010, Rotta and Noack 2011). A formal description of all combined methods is available in the online appendix (available as supplemental material at http://dx.doi.org/10.1287/mksc.2015.0950).

Finally, optimization methods such as the Louvain community detection method include a stochastic element so that, by nature, different runs of the procedure can yield (usually slightly) different solutions. These solutions are more consistent if the detected communities (or submarkets) fit well to the underlying data being analyzed. The resolution parameter impacts the number of identified communities. Cramér's V is a measure to compare the consistency of the results from multiple community detection runs and helps to find the best resolution parameter.

4.4.2. Step 2: Map Submarkets Globally. In step 2 we visualize the positions of the identified submarkets relative to one another in a single global market structure map. To do so, we select a representative for each submarket—namely, the "most central" product in each submarket, as reflected in its relationships with all products in the submarket—and subsequently map the representatives. To identify the most central product in each submarket, we again turn to network analysis techniques, using a recently developed method called harmonic centrality (Boldi and Vigna 2014). A major advantage of harmonic centrality is that it considers both direct and indirect relationships and is capable of capturing unconnected products, providing us with a more robust solution (see the online appendix for a formal description of harmonic centrality).

Since we are interested in the relative positions of entire submarkets and not just their single representatives, we must consider the relationships of all products of any given submarket to products of other submarkets. We therefore aggregate all relationships (edges) between products of different submarkets and reassign them to their respective submarket representatives, obtaining a new symmetric matrix of joint consideration consisting of submarket representatives only.

To visualize the network configuration of the identified submarket representatives, we use the Visualization of Similarities (VOS) method, which was developed in bibliometrics by van Eck and Waltman (2007). The objective of VOS is to provide a low-dimensional visualization in which objects, here products, are located in such a way that the distance between any two products reflects their similarity as accurately as possible. To this end, VOS minimizes the weighted sum of the squared Euclidean distances between all pairs of products. In our context, the more often two products were considered together (which reflects a higher similarity and thus higher weight), the greater the weight of their squared distance in the summation. See the online appendix for a formal description of the VOS method.

We chose VOS over the more traditional multidimensional scaling for four reasons. First, VOS solutions do not suffer from circular bending effects commonly associated with multidimensional scaling solutions, even when processing sparse matrices (van Eck et al. 2010). Circular bending refers to a phenomenon in which products are mapped in a circular shape or "horseshoe," leading to inaccurate interpretations of competitive relationships, since products that have weak or nonexistent competitive relationships with one another may appear closer together than they should (Kendall et al. 1970, Clark et al. 1986, Diaconis et al. 2008).

Second, VOS takes indirect similarities via third products into account to a greater extent compared with multidimensional scaling and is thus able to locate products very close to their ideal coordinates (van Eck et al. 2010). Third, whereas multidimensional scaling is prone to lumping dominant products (those that have many and strong relations to other products) together, VOS is not (van Eck et al. 2010). Lumping dominant products closely together creates inaccurate market representations in cases in which such products actually dominate individual submarkets instead of competing primarily against each other in a single submarket. Finally, VOS was designed specifically to process and map similarities based on co-occurrence data, which is how we calculated the symmetric matrix of joint consideration.

4.4.3. Step 3: Map Products Locally, Optimize Globally. In step 3 we (i) assemble a global, nonoverlapping competitive market structure map with a common scale consisting of all previously identified submarkets that (ii) also accounts for between-submarket relations of products belonging to different submarkets. Therefore, we develop a new submarket-centric mapping method that configures, positions, rotates, and dilates submarkets into a global map in three consecutive parts.

First, our submarket-centric mapping method derives a local network configuration for each submarket using VOS and then introduces all local submarket configurations into the global competitive market structure map derived in step 2. To position a (nonrepresentative) submarket member (product) in the global map, we use linear geometric transformation to calculate the product's coordinates, according to the product's distance from its submarket representative (the coordinates of all submarket representatives are set at the values they were assigned in the global map generated in step 2).

All submarkets are now in a common map but their orientation (i.e., rotation) relative to one another is random. However, since products of different submarkets can still have weak competitive relations with one another (compare step 1 in §4.4.1), submarkets must be oriented so that such relations are also accounted for. Furthermore, the ratio between distance and similarity is not necessarily the same across all submarkets since they were configured (i.e., mapped) locally. Additionally, submarkets are likely to heavily overlap in a joint space since they were originally configured locally with far fewer products to fill the map space.

We solve these problems by applying a common scale to the distances in all submarkets as well as by rotating and dilating them in such a way that between-submarket product relations are accounted for and the resulting global map configuration has little overlap.

To dilate all submarkets to a common scale, we first determine the relationship between distance (d) and disparity ($\hat{d}$) for all product relations of each submarket. Since we have similarity data (i.e., co-occurrence of products), we first monotonically transform similarity (s) to disparity ($\hat{d}$) by using the inverse of similarity ($\hat{d} = 1/s$). Furthermore, we weight individual ratios by similarity to remain consistent with the original VOS method.

Let there be C submarkets with $c \in \{1, \ldots, C\}$. Let O_c be the set of products in submarket c, and let j, k denote two products in O_c. We calculate the mean ratio ($\bar{\alpha}_c$) of distance (d_{jk}) to disparity ($\hat{d}_{jk}$) weighted by similarity (s_{jk}) as follows:

$$\bar{\alpha}_c = \frac{\sum_{i<j} (d_{ij}/\hat{d}_{ij}) \times s_{ij}}{\sum_{j<k} s_{jk}} = \frac{\sum_{j<k} d_{jk} \times s_{jk}^2}{\sum_{j<k} s_{jk}}. \tag{5}$$

In the final part of our submarket-centric mapping method, we rotate for all submarkets (C) products (O_c) of each submarket c by an angle θ_c around their submarket representative (r_c) with coordinates (x_{r_c}, y_{r_c}) and dilate all submarket members (O_c) around submarket representative (r_c) by a global scaling level (γ) divided by $\bar{\alpha}_c$ to obtain new coordinates (x'_j, y'_j) of each product j as follows:

$$x'_j = ((x_j - x_{r_c})\cos\theta_c - (y_j - y_{r_c})\sin\theta_c) \times \frac{\gamma}{\bar{\alpha}_c} + x_{r_c}, \tag{6}$$

$$y'_j = ((x_j - x_{r_c})\sin\theta_c + (y_j - y_{r_c})\cos\theta_c) \times \frac{\gamma}{\bar{\alpha}_c} + y_{r_c}. \tag{7}$$

To obtain an optimal solution, we set θ_c and γ so that the original VOS quality function is minimized

$$\text{minimize} \quad E(\mathbf{Z}; \mathbf{S}) = \sum_{j<k} s_{jk} \|\mathbf{z}_j - \mathbf{z}_k\|^2 \tag{8}$$

$$\text{subject to} \quad \|d_{jr_l}\| \le \|d_{jr_c}\|, \tag{9}$$

where $\|\cdot\|$ denotes the Euclidean norm; n: number of products; m: number of dimensions of the visualization; $\mathbf{S} = (s_{jk})$: $n \times n$ similarity matrix satisfying $s_{jk} \ge 0$, $s_{jj} = 0$, and $s_{jk} = s_{kj}$ for all $j, k \in \{1, \ldots, n\}$; $\mathbf{Z} = n \times m$ matrix containing the coordinates of the products $1, \ldots, n$; $\mathbf{z}_j$ is the jth row of $\mathbf{Z}$, containing the coordinates of product $j(x'_j$ and $y'_j)$; and $r_c \in \{r_1, r_2, \ldots, r_l, \ldots, r_C\}$.

4.4.4. Step 4: Add Asymmetry. In the final step of *DRMABS*, we introduce both global and local competitive asymmetry into our competitive market structure map.

We introduce global asymmetry, i.e., a proxy for market share, by replacing each vertex with a bubble, the size of which reflects the product's overall propensity to be included in consideration sets, given by vector **A**. The coordinates of all vertices (products) are preserved.

We introduce local competitive asymmetry by replacing each symmetric (undirected) edge with two asymmetric (directed) arcs, whose weights are the conditional probabilities corresponding to the product pair in matrix Y^*. We visualize arcs as arrows that originate in one product and point to another product so that arrows point to the competitors of those products they originate in. The heavier the weight of the arrow is, the more intense the competition.

4.5. Phase 5: Transposition of Product Attributes

Network analysis allows us to easily transpose product attributes onto a competitive market structure map by introducing vectors holding product attributes such as brand or size into the network configuration. Before doing so, we remove all information of competitive asymmetry (arcs and bubble size) as well as cluster membership (bubble color) from our map. We then visualize product attributes by varying the colors and the sizes of the bubbles representing products in our competitive market structure map. This visualization allows us to obtain insight into consumer consideration and drivers of competitive market structure.

5. Product- and Price-Comparison Site Data

Our approach is designed to process big search data stemming from any website where consumers search for and compare products in large product categories. For the purpose of this study we collected data from a product- and price-comparison site for which we developed our own tracking pixel. The ability of tracking pixels to effectively collect consumer data has raised a number of privacy concerns (Sipior et al. 2011). For our purpose, we ensured that the tracking pixel installed does not link usage behavior to personal information such as name or demographics. Moreover, as we installed the tracking pixel with the explicit consent of the product- and price-comparison site, we did not violate the terms of usage of the site.

Product- and price-comparison sites provide customers with platforms on which to search for and objectively compare various products and product offers (Giaglis et al. 2002). Such websites, which are becoming increasingly popular among consumers, actively contribute to market efficiency and reduce search costs (Ghose and Yao 2011). Some examples for product- and price-comparison sites are pricegrabber.com, idealo.de, or Google shopping.

A major advantage of using product- and price-comparison site data is that, by definition, such data span across hundreds of retailers. Thus, by using such data, we do not confine ourselves to the inventory, customer base, and sales policies (e.g., pricing and promotion) of a single retailer, eliminating the risk of modeling a competitive market structure that is artificially shaped by the specificities of a single retailer.

Furthermore, since product- and price-comparison sites generate revenue with every click on any retailer offer regardless of which product that is, they are indifferent to which products are viewed by consumers, making them an unbiased data source for product consideration.

Finally, product- and price-comparison sites capture revealed measures of consumer search at an individual level and in real time, offering insight into individual customer clickstreams, whereas other sources of online search (e.g., Google) can only provide summary information (e.g., total keyword searches). Thus, product- and price-comparison sites are a low-cost source of real-time clickstream data generated by millions of consumers searching for and comparing thousands of products in hundreds of product categories.

5.1. Structure and Mechanics of Product- and Price-Comparison Sites

The structure of product- and price-comparison sites mimics that of brick-and-mortar stores, with product categories corresponding to departments observed in physical stores. Like product- and price-comparison sites, store departments and store shelves are commonly organized according to brand, price range (e.g., premium versus "entry-level" products), and physical attributes (e.g., display size of TV sets).

Users can navigate through product- and price-comparison sites, set filters, sort lists, and enter search terms to find products. When the site finds multiple, equally well-matching products to the search terms of a user, it lists them in order of popularity by past searches of other users. Such sorting is common practice across online shops and we can even observe in brick-and-mortar stores that the most popular products are more visible to consumers than less popular products. As such, product- and price-comparison sites well represent today's market environment.

Consumers on product- and price-comparison sites engage in a detailed compensatory evaluation of products and product offers. Therefore, we can consider such sites as being situated at the end of the hierarchical model of consumer decision making proposed by Shocker et al. (1991), where consumer's product search has converged to those products they truly consider viable alternatives. The model of Shocker et al. (1991) describes the consumer's decision-making process as a sequence of phases in which the consumer initially identifies a set of alternatives, reduces the set to a smaller set based on noncompensatory rules, and then engages in a detailed compensatory evaluation of the remaining alternatives (Urban et al. 1993).

Previous studies have leveraged product- and price-comparison site data to determine whether a brand still influences consumer choices in electronic commerce (Smith and Brynjolfsson 2001), to investigate price dispersion (Baye et al. 2004, Haynes and Thompson 2008), and to estimate the determinants of clicks received by online retailers (Baye et al. 2009). We are the first, however, to use product- and price-comparison site data to observe, process, and analyze the consideration set formation of individual consumers across a large number of products to generate a visual representation of competitive market structure. Note that our data cannot be collected by Web crawlers or robots because they cannot observe search patterns of individual consumers.

5.2. Description of a Typical Consumer Clickstream

To illustrate the general structure and content of product- and price-comparison sites and to introduce the specific terminology used in this study, we provide a fictitious example of a typical consumer clickstream at such a website.

In our example, Susan needs a new smartphone and decides to search for and compare several models online. She starts by using Google to search for a smartphone from Apple. One of the first search results is a link to a popular product- and price-comparison site, which Susan clicks on. She is taken to a product-detail page for the particular Apple smartphone on the product- and price-comparison site, which we call a "page impression" for the corresponding product.

Susan is presented with a general description of the smartphone, technical details, pictures, a product video, results of several consumer tests, other consumers' opinions and ratings of the product, and a list of retailers offering the product, including links to their respective online shops. 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 product- and price-comparison site to search for "Samsung smartphone." She is presented with a list of smartphones matching her search terms on the product-listing page. Susan clicks on the Samsung smartphone she is interested in, opening its product-detail page (and generating a page impression for this particular Samsung smartphone model). Since it is late in the evening and Susan is not sure yet which smartphone she wants to buy, she decides to continue her search another time.

The following day, Susan continues her search and directly enters the URL of the product- and price-comparison site in her browser. She navigates to the smartphone product category using the category navigation menu. To reduce the list of several hundred smartphones, Susan applies filters of attributes important to her (e.g., brand, display size, WiFi capability) and sorts the results by consumer rating.

Susan clicks on a model from HTC and is taken to its product-detail page (creating a page impression for the HTC smartphone). Susan likes the HTC but wants to make sure it is the right choice for her. Therefore, she revisits the product-detail pages of the Apple and Samsung smartphones (generating another page impression for each). She studies the technical aspects of all products and, being convinced that the HTC is the best choice for her, revisits the product-detail page of the HTC smartphone (generating another page impression for it).

Having decided on the HTC, Susan 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 opens, displaying the website of the corresponding online shop. We call this a "click" on a product offer.

For the analysis proposed in this study, we would summarize the relevant information of this customer's journey as follows: Susan's consideration set size is three. Each smartphone received two page impressions. The products that Susan considered (i.e., the products in her consideration set) are referred to as being "considered jointly" with one other, such that Apple was considered jointly with Samsung, Samsung was considered jointly with HTC, and HTC was considered jointly with Apple.

6. External Validation of Big Search Data

The basic idea of our approach is to use big search data to derive consideration sets that reflect competition between products. Thus, it is essential to evaluate the quality of this kind of data, especially since online mechanisms such as product recommendations might influence consumer search but not actual purchase behavior. Therefore, we conducted an empirical study to examine the external validity of search data from a leading product- and price-comparison site. Specifically, focusing on the LED-TV market in Germany (calendar weeks 36–39 of September 2012), we analyzed the ability of the search data to predict actual market shares that were provided by GfK, the best source available in Germany for actual unit sales data by both online and off-line retailers. In addition, we collected search data from two other online data sources, namely, the leading search engine in Germany (Google) and the leading online retailer in Germany (Amazon), to compare the external validity of all three sources.

6.1. Data Collection

To obtain data from the product- and price-comparison site, we installed a tracking pixel on each page of the site. The tracking pixel enabled us to collect clickstream data of 105,606 consumers who used the site to search for and compare 1,124 products of 56 brands in the LED-TV product category. The nature of product- and price-comparison site data required us to further purify the data by excluding searches by robots or Web crawlers. Unfortunately, it is not always apparent whether a visitor is a real consumer or a robot, since robots are able to mimic consumer actions and do not identify themselves as such. However, since the mission of a robot is to systematically collect information that is too cumbersome to collect by hand, we assume that it is unlikely for a real consumer to search for and compare a very large number of products. Our collected data show that 99.9% of all captured clickstreams span across 16 products or fewer. We therefore chose 16 products as a cut-off level and eliminated all clickstreams containing more than 16 products. Finally, we determined the frequency of consumer consideration for each of our 1,124 products (compare with §4.3.1).

We used Google's keyword search tool to obtain the monthly volume of Google keyword searches for LED-TVs. We collected the keyword search volume for 914 LED-TVs using individual product names (e.g., KDL-40HX755) as search terms.

Using Amazon's product advertising API (application programming interface), we polled the list of "also viewed products" displayed with each product on Amazon's website on a daily basis in September 2012. These data enabled us to identify the number of instances in which a given LED-TV was included in the "also viewed" lists of other LED-TVs. Overall, we collected data on 964 LED-TVs using Amazon's API.

Finally, we obtained actual LED-TV unit sales data from GfK for 2,003 LED-TVs. GfK aggregates total sales volume (units sold) in a few popular product categories, among them LED-TVs, on a weekly basis from a retailer panel covering 93% of the market. Retailers, including large discounters, report both online and in-store sales to the market researcher, who aggregates the data and creates sales reports, which are subsequently made available to all retailers participating in the panel. Note that since GfK's sales reports usually span longer periods of time, the September 2012 report includes many products that are no longer available in the market.

6.2. Matching Products Across Data Sources

A comparative validation of multiple data sources requires a common sample of products. Since each data source has its own identifier (for Google it is simply the product name as a search term), we must first match products across data sources. Moreover, not all search data sources supply data on all products available in the market. Finally, GfK's listing also includes products that are no longer available in the market and generate no sales at all.

Generally, products are identified at the SKU level (stockkeeping unit) using EANs (International Article Numbers). However, EANs turn out to be a poor key for matching products, since in many cases, a single product has multiple EANs, with some supplied from one data source and others supplied from another data source. The reason for multiple EANs are manifold and include special or new packaging as well as the assignment of multiple EANs to identify where a product was originally shipped by the manufacturer.

We therefore matched products by hand, comparing product numbers and looking up individual products online when no clear match could be made. For this cumbersome but necessary endeavor to validate our data, two independent researchers spent close to 100 hours matching products across all four data sources by comparing product names, product numbers, technical specifications, and product pictures. The matched lists were then compared, and a joint rematching was carried out in cases of discrepancies. We ultimately obtained a common sample of 549 exact product matches across all sources.

6.3. Findings of Data Validation

We tested the data from each source against the total unit sales volume (online and in-store) from GfK's retailer panel using Pearson's two-tailed correlation analysis (see Table 5). The highest correlation with sales data was obtained with the consideration set data from the product- and price-comparison site ($r = 0.753$ across all products), followed by Google's keyword search volume ($r = 0.698$) and finally Amazon's aggregated "also viewed" data ($r = 0.347$).

The strong positive correlation of the consideration set data with actual unit sales confirms our notion that users of product- and price-comparison sites are in a late stage of the consumer decision-making process (Shocker et al. 1991), suggesting that these consumers are relatively close to purchase.

Google's keyword search volume for product names also correlated strongly ($r = 0.698$) with unit sales. This finding is not surprising, since consumers who search for a precise product code such as "KDL-40HX755" might well be looking for an online retailer from which to buy the product or looking for online documentation and troubleshooting guidance after just having purchased the product.

Finally, Amazon's aggregated "also viewed" data did not correlate as strongly ($r = 0.347$) with actual unit

Table 5 External Validation of Online Search Data

		Retailer panel *total unit sales*	Product and price comparison *frequency of consideration*	Google *search volume*	Amazon *frequency on also viewed lists*
All products, $N = 549$	*Retailer panel total unit sales*	1.000**	0.753**	0.698**	0.347**
Small display <33", $N = 204$	*Retailer panel total unit sales*	1.000**	0.750**	0.701**	0.511**
Medium display 33"–45", $N = 174$	*Retailer panel total unit sales*	1.000**	0.835**	0.743**	0.242**
Large display >45", $N = 171$	*Retailer panel total unit sales*	1.000**	0.812**	0.805**	0.311**

Note. Two-tailed Pearson.
**Significant at 0.01 level.

sales as did search data from the other two data sources. This lower correlation might be attributed to the fact that Amazon's "also viewed" lists are based on previously aggregated search data that have been truncated and potentially further processed or normalized.

To check the robustness of our results, we reran our correlation analysis on three approximately equal-sized subgroups of different display sizes, namely, small, medium, and large LED-TVs. We find stable results across all subgroups (see Table 5), with data from the product- and price-comparison site performing consistently better than Google and Amazon data.

The strong correlation between search data from the product- and price-comparison site and actual market outcome suggests that these data are an excellent source for conducting competitive analysis based on consumer consideration sets.

7. Empirical Application of our Approach

The aim of this empirical study is to demonstrate how our approach can be used to analyze and visualize asymmetric competition in a large market consisting of hundreds of products. Specifically, we aim to generate a visual representation of competitive market structure, identify submarkets, visualize asymmetric competition, transpose product attributes onto the competitive market structure map, and zoom in on a single submarket to get a better understanding of the local competitive situation in that submarket. Finally, we aim to compare our solution to solutions obtained with traditional mapping methods such as multidimensional scaling.

We chose the LED-TV market for this empirical study for three reasons. First, LED-TVs have become highly popular consumer durables, with global sales expected to exceed 190 million units in 2013 (Gagnon and Torii 2012), making LED-TVs an important market for many manufacturers and retailers. Second, since the LED-TV market consists of over 1,000 products, the use of traditional approaches to identify and visualize competitive relations among these products would be very cumbersome if not impossible. Third, since the LED-TV market is well known with easy-to-understand product attributes, it is relatively easy for us to test our findings for face validity.

7.1. General Findings in the LED-TV Market

In phase 1 of our approach, we construct 105,606 consideration sets from the collected clickstream data (compare with §6.1) and generate the symmetric matrix Y for 1,124 products, capturing all $N \times (N-1)2 = 631{,}126$ possible relationships between pairs of products.

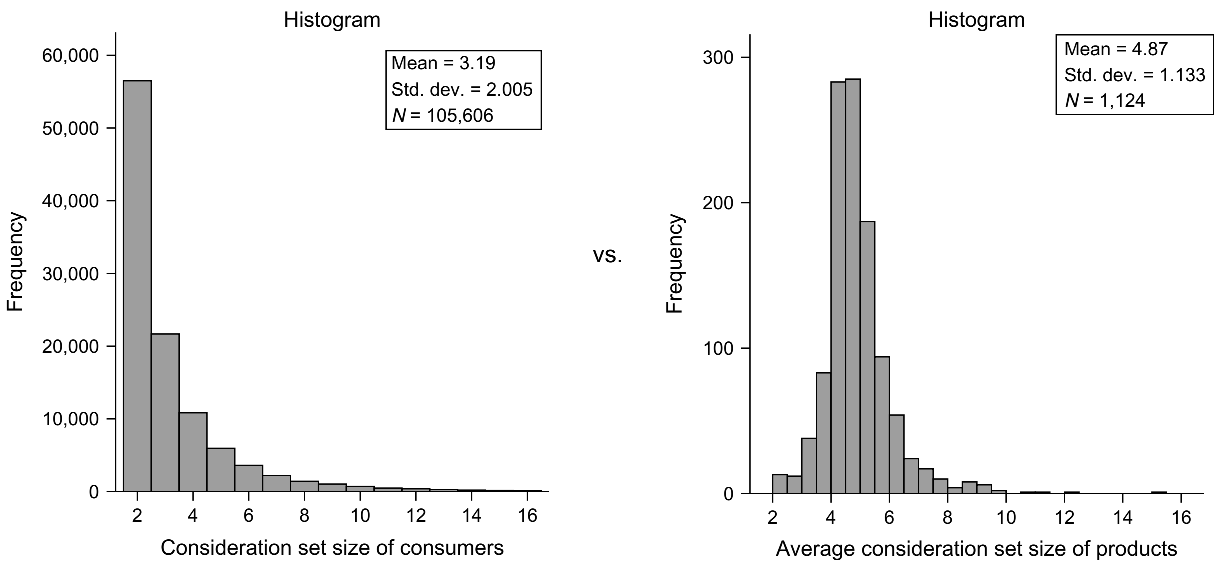
We find a mean consumer consideration set size of 3.19 with a standard deviation of 2.005 and a positively skewed distribution (see Figure 4). This finding is in line with previous research (e.g., Aurier et al. 2000, DeSarbo et al. 2006), which, using survey data, found consumer consideration sets to contain two to six products. On the other hand, more recently, Kim et al. (2010) estimated the mean consideration set size of consumers to be 14. The difference in these findings might be attributed to the nature of the collected data. Whereas Aurier et al. (2000) and DeSarbo et al. (2006) collected individual consumer consideration sets (as we do), Kim et al. (2010) collected aggregate data on consumer consideration.

When we aggregate consideration sets at the product level, we find a mean product consideration set size of 4.87 with a standard deviation of 1.133 (see Figure 4). A closer look suggests that there are some products that have few substitutes, as they are jointly considered with very few other products. On the other hand, there are other products that are jointly considered with many products, pointing to more competition with less differentiation.

Table 6 depicts the top 10 brands and products by share of consumer consideration, based on vector **A**, which holds the number of consideration sets in which each product was included.

Samsung clearly dominates the market with a share of nearly 43%. Overall, the top 10 brands clearly dominate the LED-TV market with a joint share of 96.85%. The remaining 46 brands make up only 3.15% of the

Figure 4 **Consideration Set Size of Consumers vs. Average Consideration Set Size of Products**



market. The HHI of 4,529 confirms that the LED-TV market is highly concentrated at the brand level.

At the product level, Samsung again dominates, supplying nine out of the top 10 products in the LED-TV market. Yet, with 1,124 products competing, the LED-TV market has a very low market concentration at the product level, as indicated by an HHI of only 50.

7.2. Visualization of Asymmetric Competitive Market Structure

Figure 5 displays the asymmetric competitive market structure map for 1,124 LED-TVs with minimal overlaps. Submarket membership can be easily discerned both by bubble color and by the spatial mapping of products into groups; the latter becomes crucial to interpreting competitive market structure when we later use color coding of bubbles to transpose product attributes onto the competitive market structure map instead of using colors to identify submarkets (see Figure 6). The Louvain community detection algorithm with the addition of a resolution parameter and enhanced by multilevel coarsening and refinement identifies 30 distinct submarkets. The most consistent solutions are obtained with a resolution parameter of 3.1, as indicated by a Cramér's V of 0.962.

We use bubble size to indicate global competitive asymmetry (i.e., the frequency of consideration as a proxy for market share), which makes it easy to identify dominating products in the market. Local asymmetry (i.e., conditional probability), as expressed by the arrows connecting competing products, can best be seen by zooming in on areas of the asymmetric competitive market map (see §7.4 for an example). The presence of both very heavy as well as very light arrows indicates high degrees of local competitive asymmetry.

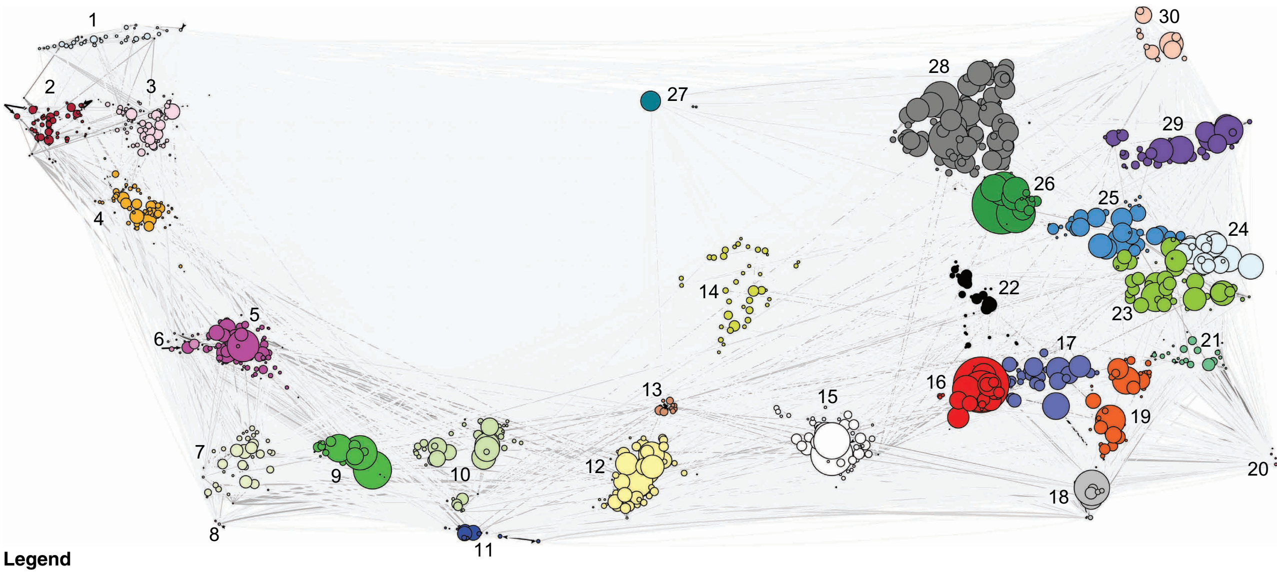
As shown in Figure 5, LED-TVs are clearly organized into submarkets such that within-submarket competitive relations among products are stronger than between-submarket competitive relations. Furthermore, submarkets whose products have stronger between-submarket competitive relations are located closer to one another. Likewise, these submarkets are oriented such that individual products (in different submarkets) that compete more strongly with one another are positioned closer together. We observe an instance of an especially strong local competitive asymmetry in submarket 2 (Figure 5) where there are two very heavy arrows originating in a tiny bubbled product (Orion TV24LB860) at the very edge of the map and pointing at two larger bubbled products (Orion 24LB890 and Telefunken T24EP970CT) more toward the center of submarket 2. This constellation indicates

Table 6 **Market Share According to Consumer Consideration**

Top 10 brands		Top 10 products	
Brand	Share of 56 brands	Product	Share of 1,124 products
Samsung	42.61%	Samsung UE46ES6300	2.91%
Philips	15.93%	Samsung UE40ES6300	2.59%
LG	14.36%	Samsung UE40ES5700	1.47%
Panasonic	8.07%	Samsung UE40ES6710	1.39%
Sony	6.30%	Samsung UE55ES6300	1.31%
Toshiba	4.07%	Samsung UE46ES6710	1.26%
Sharp	2.28%	Samsung UE32ES6300	1.21%
Grundig	1.99%	Samsung UE40EH5200	1.11%
Loewe	0.83%	Samsung UE55ES8090	1.10%
Telefunken	0.42%	Philips 40PFL5507K	1.05%
Other	3.15%	Other	84.59%
HHI	4,529	HHI	50

Note. HHI, Herfindahl–Hirschman index.

Figure 5 Visualization of Asymmetric Competitive Market Structure Map of 1,124 LED-TVs



Legend
Bubbles represent individual products (SKUs)
Bubble color indicates submarket membership
Bubble size indicates global competitive asymmetry (consideration frequency)
Arrows represent local competitive asymmetry and point at competitors of the product they originate in
Arrow weight indicates how intense a competitive relationship is: the darker and thicker the arrow, the more intense the relationship
Submarkets are numbered 1 through 30

that Orion 24LB890 and Telefunken T24EP970CT are always present in the consideration sets where Orion TV24LB860 is present as given by a conditional probability of 1. The returning arrows, however, are so light (reflecting conditional probabilities close to 0) that they are invisible without zooming in further on the map. These invisible arrows indicate that Orion TV24LB860 is hardly ever present in the consideration sets in which Orion 24LB890 and Telefunken T24EP970CT are present.

7.3. Using Product Attributes to Better Understand Competitive Market Structure

We derive competitive market structure solely from the clickstreams of consumers searching for and comparing products. We now transpose product attributes onto our competitive market structure map. Our aims are threefold: first, we aim to better understand how the LED-TV market is organized; second, we aim to test our findings for face validity; and third, we aim to determine whether it is necessary to look at the entire market or if a smaller, a priori defined market based on key product features is sufficient.

The attributes we selected for this purpose are brand, display size, and 3D capability, although additional attributes could easily be included.

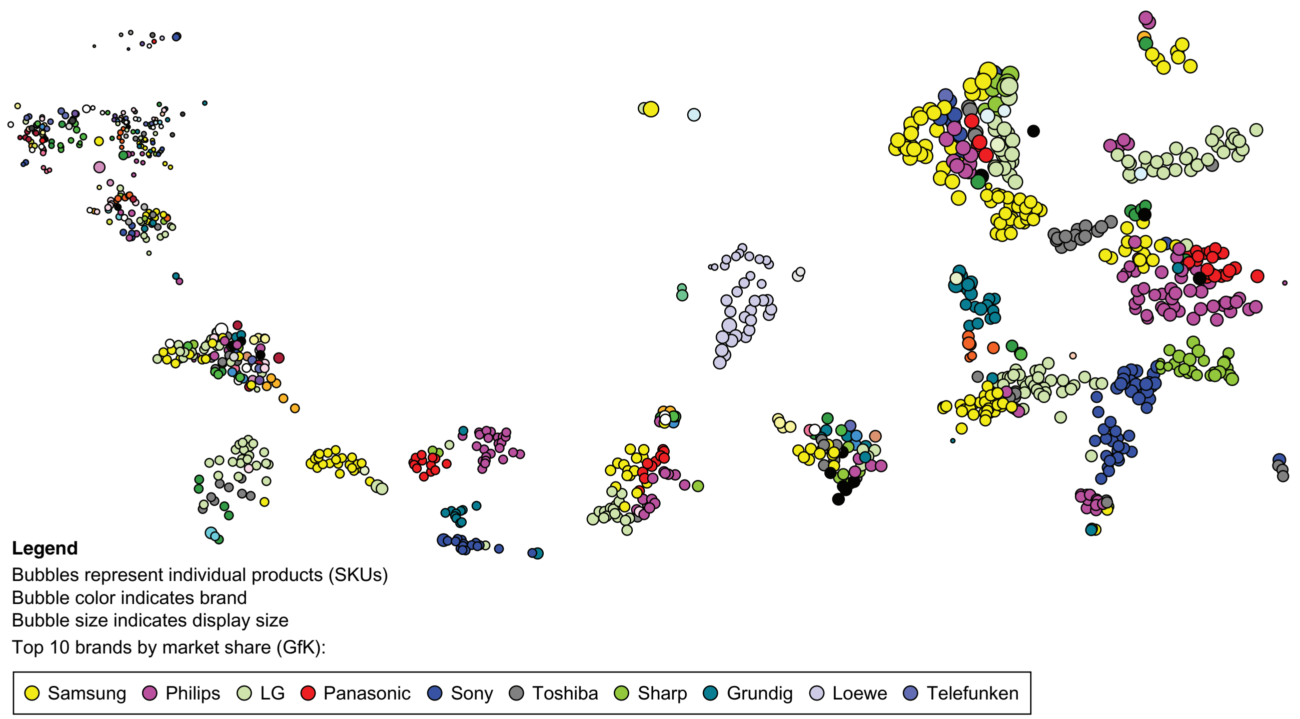
For each product j, the product attributes brand, display size, and 3D capability are stored in vectors $\mathbf{B}$, $\mathbf{D}$, and $\mathbf{E}$, respectively, and these vectors are then incorporated into our network configuration. To better visualize combinations of product attributes, we do not display competitive asymmetry information (represented in Figure 5 by bubble size and arrow weight) or submarket memberships (color) when displaying product attributes.

Given the heavy branding efforts of manufacturers in this highly competitive market as well as the need of consumers to select a television that physically fits best into their living rooms, we intuitively expect brand and display size to drive the organization of our competitive market structure map to some extent. Such a finding, although not associated with statistical inference methods, would support the face validity of our approach. A perfect fit of brand and display size to the organization of our competitive market structure map, however, would cast doubt on the value of our new approach, since managers could simply use predominant attributes to identify competing products.

Figure 6 depicts a competitive market structure with vector $\mathbf{B}$ (brand) and vector $\mathbf{D}$ (display size) transposed onto product coordinates. Bubble color indicates brand, and bubble size indicates display size. We find a number of same-brand products clustered tightly together, as indicated by bubble color. Overall, brand seems to contribute to the organization of our competitive market structure map of the LED-TV market, as many submarkets are made up of only a few brands.

Display size appears to be another driver for market structure, as we find small displays concentrated

Figure 6 Using Brand and Display Size to Understand Competitive Market Structure of 1,124 LED-TVs



toward the upper left of the map, with increasing display size as we move toward the right (see Figure 6). Note that the products with larger displays (top right) are predominantly offered by leading brands such as Samsung, LG, and Philips, whereas small displays (top left) are offered by a very large number of smaller brands. Property fitting of product coordinates to display size indicates a fair but not perfect fit ($R^2 = 0.688$). Given that, at the time of this empirical study, the trend in the LED-TV market was toward developing larger and larger displays, we would expect the industry leaders to dominate this area of the market, which supports the face validity of our solution.

Finally, we transpose a new and innovative product attribute, 3D capability, onto our competitive market structure map to determine whether it is possible to use such an attribute for an a priori definition of smaller submarkets that can be analyzed with traditional mapping methods. A well-defined submarket (or group of contiguous submarkets) consisting of 3D LED-TVs would lend support to such an a priori market definition.

We transpose 3D capability (vector **E**) onto our competitive market structure map using red triangles in Figure 7. Clearly, 3D capability is not a submarket-defining feature, since 3D LED-TVs are scattered across most submarkets. Consequently, an a priori market definition of 3D LED-TVs would have led to a wrong competitive market structure representation.

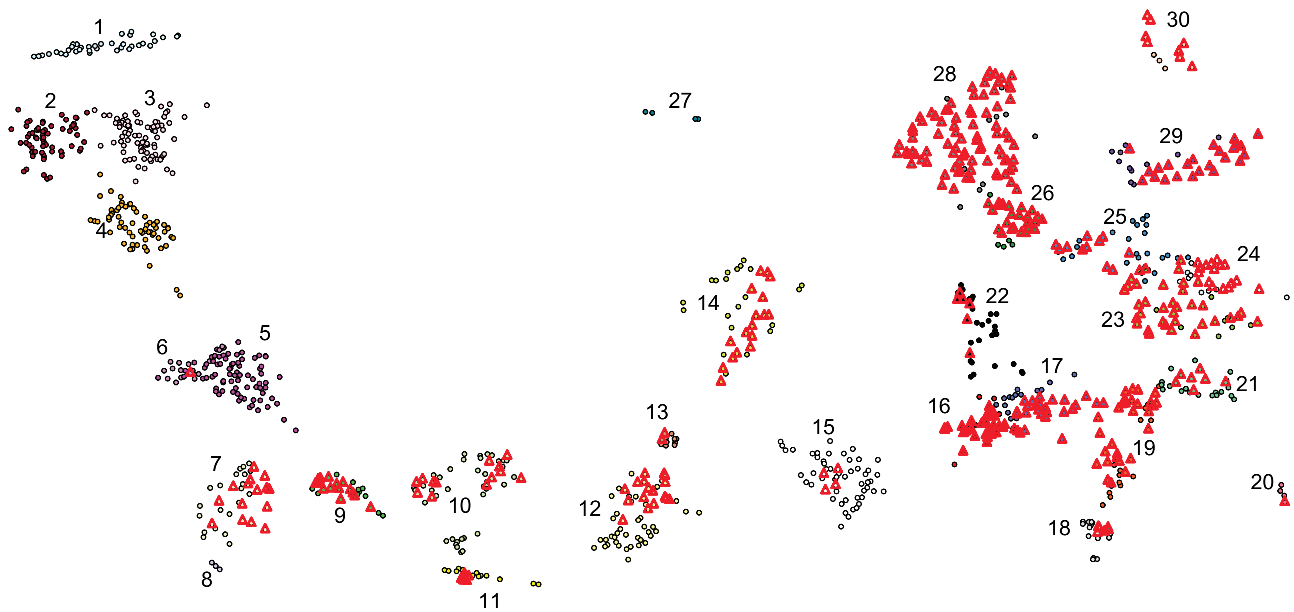
Evidently, a major advantage of a single map that comprises the entire market is that new products with new features (e.g., 3D TVs or curved displays) are naturally positioned into either existing or into newly formed submarkets based directly on consumer perception. This posterior market definition eliminates the uncertainty of whether a new product feature will lead to the formation of a new submarket or not, which we believe is an inherent problem when submarkets are defined a priori.

7.4. Zooming in on Submarkets

Although a complete asymmetric competitive market structure map of 1,124 products helps us to understand the overall market, it is too coarse to reveal individual competitive relations among products. Since our approach allows us to zoom in on any area of the competitive market structure map, we now take a closer look at an individual submarket. For the purpose of this empirical study we chose submarket number 19 (represented by orange bubbles in Figure 5), which consists entirely of Sony products.

To obtain an even more fine-grained segmentation, we apply the Louvain community detection method described in §4.4.1 again to just this submarket. We identify three fine-grained segments, which we denote by color in Figure 8. A closer look at Sony's product codes, used as labels in Figure 8, reveals that the yellow segment at the top right consists entirely of 46-inch products, and the red and green segments consist of 40-inch products (as well as a single 42-inch product). The division of 40-inch TVs into the red and green segments can be explained by the product series as

Figure 7 **3D TVs in the Competitive Market Structure Map of 1,124 LED-TVs**



well as the street prices of the products included in each submarket. With the exception of one product (the KDL-40HX805), all 800-series products are in the lower, green segment. Higher series numbers for Sony TVs indicate that they have more and better features. Furthermore, the mean street price of the red segment is €733, and that of the green segment is €1,520. Since the KDL-40HX805 has a street price of €900, it is logical that it competes more with products in the red segment.

From our data set we know that Sony's strongest product is its 40-inch KDL-40HX755 product, ranked 15th overall with a market share of 0.75%. We would therefore expect the Sony KDL-40HX755 to be positioned toward the center of Figure 8, to have the largest bubble size (global competitive asymmetry, as a proxy for market share) and to have numerous heavy arrows pointing at it (local competitive asymmetry), which is clearly the case. Beyond this top Sony product we also find several other products that are strong in terms of global and local competitive asymmetry (e.g., KDL-46HX755, KDL-40HX855, KDL-46HX855) relative to the other Sony products of this submarket. On the downside, there are also numerous Sony products with very small bubbles (meaning that they are not heavy global competitors since they only receive a low consideration) and heavy arrows pointing to other Sony products, indicating them as their strongest competitors.

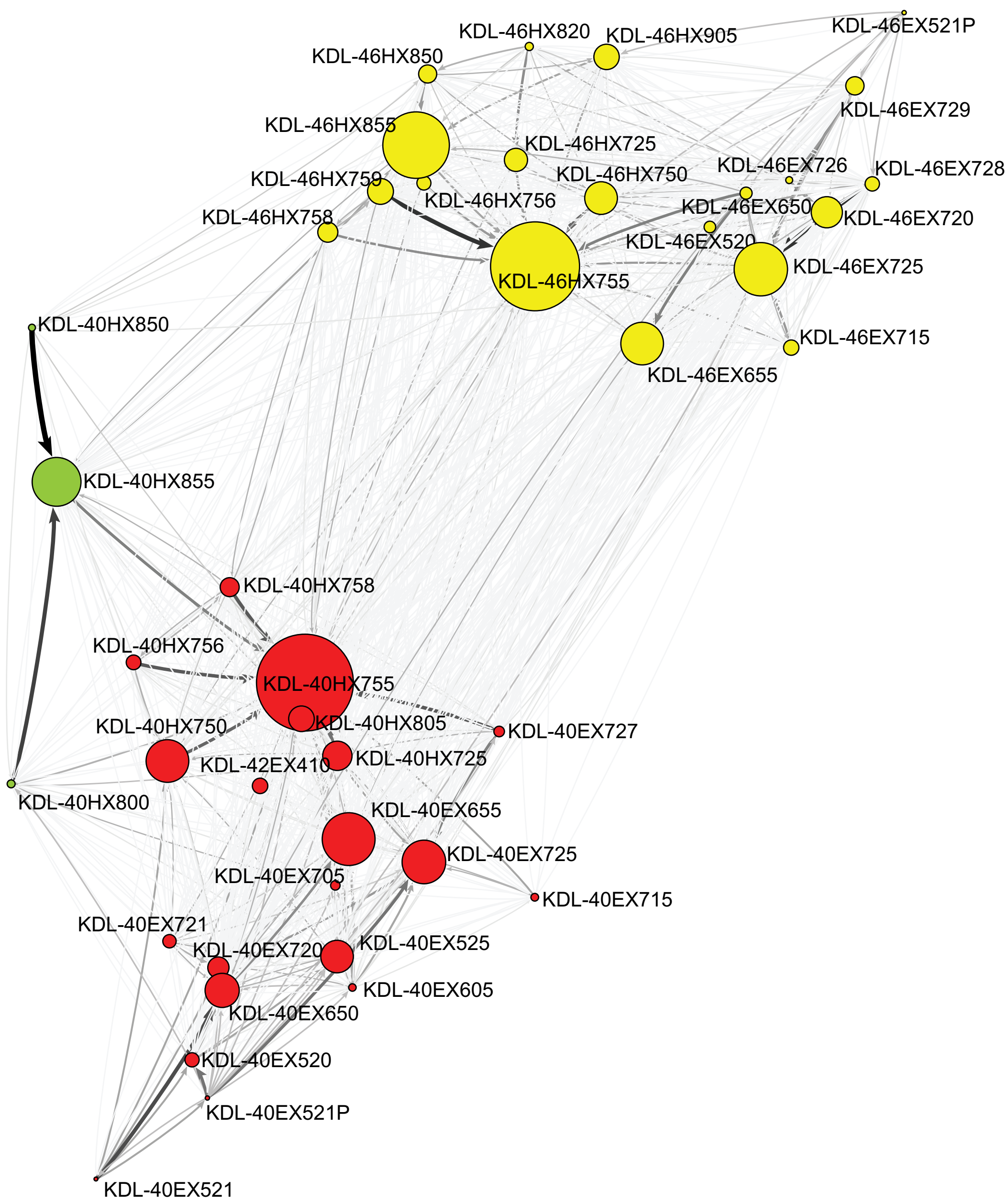
Overall, we find that the spatial division of products into three areas in Figure 8 is indeed supported by the identification of three fine-grained market segments, providing evidence that the segmentation and mapping methods combined in *DRMABS* are consistent with one another. Moreover, the way competitive asymmetry, display size, product series, and street price are recovered by *DRMABS* in Figure 8 provides support to the face validity of our approach. Still, although we provided several tests for evaluating the validity of our approach, we recognize that these tests are descriptive and not based on any statistical inference methods, which prevents us from quantifying any sampling error.

7.5. Managerial Insights

Our new approach provides both manufacturers and retailers with insights that they cannot obtain from other sources such as market share or sales reports. Manufacturers can use our asymmetric competitive market structure maps to quickly see how a market is organized, how many submarkets exist, which competitors they face in each submarket, and how strong these competitors are. Retailers, on the other hand, can use our new approach to make better purchasing and inventory management decisions because they probably want to cover many segments without stocking too many products.

Specifically, manufactures can learn from Figure 5 that most LED-TV submarkets are dominated by a few products, as indicated by these products' large bubble sizes (where bubble size captures global competitive

Figure 8 **Zoom-in on Submarket Number 19 (All Products Are from Sony)**



Legend
Bubbles represent individual products
Bubble color indicates cluster membership
Bubble size indicates global competitive asymmetry (consideration frequency)
Arrows represent local competitive asymmetry and point at competitors of the product they originate in
Arrow weight indicates how relevant a competitor is: the darker and thicker, the more relevant the competitor
Products labeled with model name (KDL indicates a Sony product with two digits after hyphen indicating its display size)

asymmetry). In fact, the top 10 LED-TVs (compare with Table 6) do not compete primarily against one another, but rather against products in their respective submarkets, which is an insight that managers cannot attain solely by considering the products' market shares.

Furthermore, our maps reveal that a given brand might face different competitors in different areas of the market, and they enable manufacturers to observe who these competitors are. For instance, Sony's (dark blue) closest competitors in the area of 34″ to 37″ TVs (bottom left of Figure 6) are Grundig (dark cyan) and Panasonic (red). However, in the 40″ to 46″ area of the market (bottom right of Figure 6), Sony faces different competitors, namely, LG (lime), Sharp (green), and Philips (magenta). Consequently, product line managers must align their targeting, positioning, and communication strategies to the specific competitors they face in a specific area of the market, especially when competitors have different strengths and follow different strategies. Note that the orientation of individual submarkets relative to one another is crucial in correctly assessing who the closest competitors in nearby submarkets are. Since our model *DRMABS* accounts for between-submarket relations by rotating submarkets, we are able to provide insights beyond a mere series of individual submarket maps.

By taking a closer look at maps such as the one in Figure 6, manufacturers can attain insight into the positioning strategies of different brands. For instance, whereas Samsung products are present across the entire market, products of the premium brand Loewe are concentrated in one central submarket. Although Loewe's managers may consider it to be good news that the brand practically defines its own submarket, our map also shows that the "Loewe submarket" is isolated from other submarkets and draws relatively little consumer consideration (compare with Figure 6). This insight should alarm Loewe managers since it essentially means that the once highly popular Loewe brand is dropping out of consumers' consideration sets. Indeed, in line with this troubling insight, Loewe filed bankruptcy in 2013, only one year after our data collection.

Similarly, managers of Sony can learn from Figure 8 that the large number of products and the presence of heavy competitive asymmetry within submarket 19 (a Sony-exclusive submarket) point to a potential within-brand cannibalization problem.

Retailers can use our maps for guidance in selecting and managing their product inventory. Most retailers have both budget and space constraints when stocking products for sale. Within these constraints they must decide which products to order from more than 1,000 products offered by manufacturers. Wrong decisions can leave them either with overstock that does not sell, or shortages of "hot products" that prevent them from meeting the demand of their customers. In both cases they end up losing money. Our new approach informs retailers in a timely and low-cost way which products are most popular in each submarket. Retailers can thus easily serve a broad spectrum of consumer needs with a relatively small number of products (e.g., 30 LED-TVs if they select the most popular one in each submarket). Furthermore, retailers can obtain an indication of how great the overall market demand for each product is (global asymmetry) and balance order quantities accordingly. Retailers wishing to offer some alternatives to any given product can use our map to find the respective substitutes.

7.6. Model Comparison

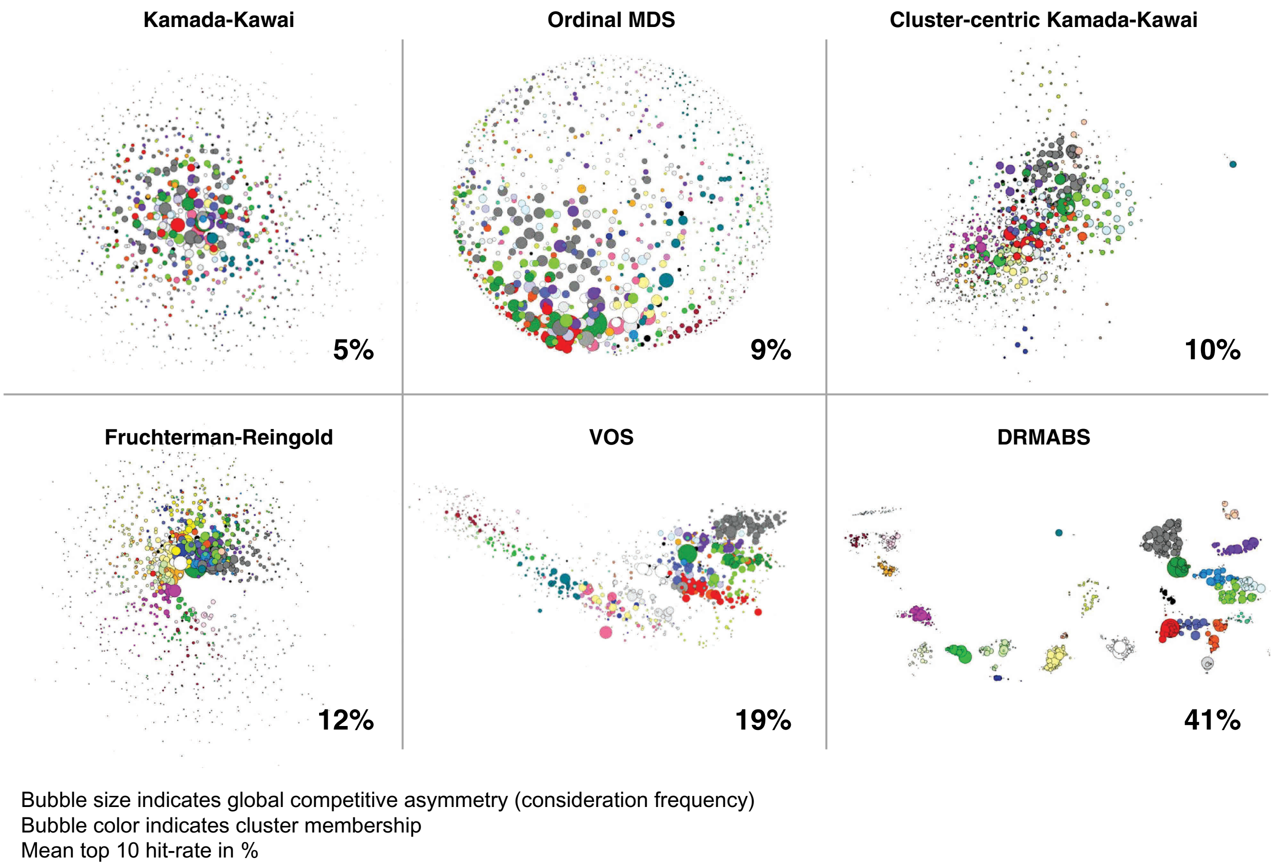
We now compare our model empirically with traditional models for competitive market structure mapping based on our matrix Y_{jk} of joint product consideration. Specifically, we compare our model to solutions using ordinal multidimensional scaling (i.e., PROXSCAL), Kamada and Kawai (1989) and Fruchterman and Reingold (1991) force-directed drawing, plain VOS mapping (van Eck and Waltman 2007), and a cluster-centric version of Kamada–Kawai's force-directed drawing (De Nooy et al. 2011).

To determine which model performs best, we proceed in two steps. First, we visually inspect each mapping solution for potential weaknesses such as circular bending, lumping of dominant products, and poor submarket recovery. For this purpose, in each map, we use bubble size to indicate global competitive asymmetry as a proxy for market share and bubble color to indicate submarket membership.

Second, we evaluate the performance of each model. Since each model optimizes its own quality metric, we must first define a common quality metric across all models. Faure and Natter (2010) show that the quality of a mapping solution can be measured with different metrics whose selection depends on the objective of the mapping solution. Since it is our objective to map the strongest competitors of any product as close to the product as possible, we propose that a hit-rate indicating how many of the strongest competitors in terms of joint consideration are among the closest products in terms of distance is an appropriate quality metric. Therefore, for each map solution, we calculate each product's top 10 hit-rate (i.e., the share of the product's top 10 competitors that are included among the 10 products located closest to it on the map), and compute the average top 10 hit-rate across all products. This hit-rate enables us to determine which method produces a map that best fits the underlying competition (similarity) data.

The results of our model comparison are depicted in Figure 9. We observe both circular bending and heavy lumping of dominant products in the solutions of Kamada–Kawai, Fruchterman–Reingold, and ordinal multidimensional scaling. Using VOS alone leads to a mapping solution whose shape and general submarket positions resemble that of our model's mapping solution, but the hit-rate of VOS is less than half as high (19%). The cluster-centric Kamada–Kawai solution, which does not optimize submarket rotation and dilation, suffers from heavy overlapping of submarkets. *DRMABS* clearly outperforms all other methods in terms of hit-rate (41%), does not suffer from submarket

Figure 9 **Comparison of New Model *DRMABS* with Other Models That Display the Competitive Structure of a Market Comprising 1,124 LED-TVs**



overlaps, and does not exhibit circular bending or lumping of dominant products.

Finally, note that in VOS, the sum of many very weak relationships to products outside a submarket can lead to a delusion of the much stronger relationships to products within the submarket, resulting in a "blurred" mapping solution where products are "pulled" away from their closest competitors by a larger number of very weak competitors. The much higher hit-rate of *DRMABS* (41%) nicely shows the effect of such "delusion" when comparing VOS to *DRMABS*. In addition, *DRMABS* also depicts competitive relationships for products of different submarkets using arrows of local competitive asymmetry. Using these arrows, users of *DRMABS* can easily check whether there are any notable competitors of a given product in other (nearby) submarkets and which ones they are.

8. Summary and Conclusions

The stream of literature devoted to competitive analysis documents the importance of understanding competitive relationships (e.g., Erdem 1996, Bergen and Peteraf 2002, DeSarbo et al. 2006). Yet researchers currently have limited ability to comprehensively analyze competitive relationships in large markets, which are becoming increasingly prevalent—a typical durable-goods market might consist of hundreds of products linked to several dozens of brands. We extend the line of competitive analysis research by developing a new approach that, using clickstream data, creates a nonoverlapping visualization of the competitive market structure in a large durable-goods category containing over 1,000 products. Furthermore, our visualization method provides insight into both global and local levels of competitive asymmetry.

Our approach uses a new data source (product- and price-comparison sites) and includes a newly developed model called *DRMABS* (*D*ecomposition and *R*eassembly of *MA*rkets *B*y *S*egmentation) that combines methods from multiple research disciplines with a newly developed submarket-centric mapping method. The proposed methodology produces easy-to-interpret and face-valid maps of asymmetric competitive market structure for large markets consisting of more than 1,000 products. The maps are designed to prevent products from overlapping or being lumped together, thus enabling submarkets to be easily identified.

In three empirical studies we have (i) demonstrated the need to analyze big data in large consumer durable markets, (ii) externally validated our data, and (iii) applied our approach to the LED-TV market and compared our model *DRMABS* to existing models. Our first empirical study shows that analysis limited to only a few products (i.e., the top 50 products of a large market) would exclude crucial information from the analysis (in the products we examine, almost 50% of the sold products and 77% of the competing brands would be excluded). These results highlight the importance of analyzing big data to obtain a comprehensive view of the competitive structure of a large market.

Our second empirical study is the first in marketing to test and compare the external validity of search data from several online data sources using actual unit sales data obtained from a leading market research institute. We find that search data from a product- and price-comparison site exhibits higher external validity compared with search data from Google and Amazon.

In our third empirical study we demonstrated our new approach in an application to the LED-TV market. We collected and processed the clickstreams of over 100,000 consumers from a product- and price-comparison site and used *DRMABS* to generate an asymmetric competitive market structure map consisting of 1,124 products of 56 brands. By transposing easy-to-understand product attributes onto our map (i.e., brand, display size, and 3D capability), we generated new insights for manufacturers and retailers, confirmed the face-validity of our maps, and demonstrated that a priori market definitions based on product attributes can lead to wrong representations of competitive market structure. Finally, we showed that *DRMABS* outperforms traditional mapping models.

The approach proposed herein for modeling and mapping competitive market structures thus lays an important foundation for the analysis of today's large markets. At the same time, it creates a number of opportunities for future research. Potential extensions include enriching the data analyzed or expansion of the methods employed. For instance, future maps might incorporate additional data on product attributes and consumer preferences, which could introduce latent dimensions of consumer satisfaction with product attribute quality (Tirunillai and Tellis 2014). Additionally, a partitioning of consumers could be introduced to specifically account for consumer heterogeneity (Vera et al. 2009). Furthermore, recent work in the area of submarket detection such as that of France and Ghose (2016) could be integrated into *DRMABS* to test for submarket structures using sales and conversion data.

Finally, our approach can easily be automated, since it requires no manual intervention during data collection, processing, and map generation. Such automation enables decision makers to conduct time-series analysis as a basis for tracking the impact of new product launches or communication campaigns.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mksc.2015.0950.

Acknowledgments

The authors would like to thank Arvind Rangaswamy, Friedrich Fleischmann, Jochen Rainer, Martin Natter, Martin Sinner, Nadia Abou Nabout, Bettina Steinbrenner, Thomas Otter, Ludo Waltman, the marketing departments at BI Norwegian Business School, Cambridge University, and Tilburg University as well as the Chairs of the EMAC 27th Doctoral Colloquium and last but not least the editors, the associate editor, and reviewers for their valuable feedback and support that helped continuously improve this manuscript.

References

Allenby GM, Rossi PE (1991) Quality perceptions and asymmetric switching between brands. *Marketing Sci.* 10(3):185–204.

Aurier P, Jean S, Zaichkowsky JL (2000) Consideration set size and familiarity with usage context. *Adv. Consumer Res.* 27(1):307–313.

Baye MR, Morgan J, Scholten P (2004) Price dispersion in the small and in the large: Evidence from an Internet price comparison site. *J. Indust. Econom.* 52(4):463–496.

Baye MR, Gatti JRJ, Kattuman P, Morgan J (2009) Clicks, discontinuities, and firm demand online. *J. Econom. Management Strategy* 18(4):935–975.

Bergen M, Peteraf MA (2002) Competitor identification and competitor analysis: A broad-based managerial approach. *Managerial Decision Econom.* 23(4–5):157–169.

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J. Statist. Mechanics: Theory Experiment* 2008(10):1742–1756.

Boldi P, Vigna S (2014) Axioms for centrality. *Internet Math.* 10(3–4): 222–262.

Buja A, Swayne DF, Littman ML, Dean N, Hofmann H, Chen L (2008) Data visualization with multidimensional scaling. *J. Comput. Graphical Statist.* 17(2):444–472.

Clark WC, Carroll JD, Yang JC, Janal MN (1986) Multidimensional scaling reveals two dimensions of thermal pain. *J. Experiment. Psych.: Human Perception Performance* 12(1):103–107.

De Nooy W, Mrvar A, Batagelj V (2011) *Exploratory Social Network Analysis with Pajek* (Cambridge University Press, New York).

DeSarbo W, Jedidi K (1995) The spatial representation of heterogeneous consideration sets. *Marketing Sci.* 14(3):326–342.

DeSarbo WS, Grewal R (2007) An alternative efficient representation of demand-based competitive asymmetry. *Strategic Management J.* 28(7):755–766.

DeSarbo WS, Grewal R, Wind J (2006) Who competes with whom? A demand-based perspective for identifying and representing asymmetric competition. *Strategic Management J.* 27(2):101–129.

DeSarbo WS, Manrai AK, Manrai LA (1993) Non-spatial tree models for the assessment of comparative market structure: An integrated review of the marketing and psychometric literature. Eliashberg J, Lilien G, eds. *Marketing*, Handbook Oper. Res. Management Sci. (North-Holland, Amsterdam), 193–257.

Diaconis P, Goel S, Holmes S (2008) Horseshoes in multidimensional scaling and local kernel methods. *Ann. Appl. Statist.* 2(3):777–807.

Erdem T (1996) A dynamic analysis of market structure based on panel data. *Marketing Sci.* 15(4):359–378.

Error B, Error C (2004) Efficient click-stream data collection. US Patent US 20040098229 A1 (File date, June 26, 2003).

Faure C, Natter M (2010) New metrics for evaluating preference maps. *Internat. J. Res. Marketing* 27(3):261–270.

Fortunato S (2010) Community detection in graphs. *Physics Reports* 486(3):75–174.

Fortunato S, Barthelemy M (2007) Resolution limit in community detection. *Proc. Natl. Acad. Sci. USA* 104(1):36–41.

France S, Ghose S (2016) An analysis and visualization methodology for identifying and testing market structure. *Marketing Sci.* 35(1):182–197.

Fruchterman TMJ, Reingold EM (1991) Graph drawing by force-directed placement. *Software: Practice Experience* 21(11):1129–1164.

Gagnon P, Torii H (2012) NPD display search quarterly global TV shipment and forecast report. NPD Display Search, Englewood, CO.

Ghose A, Yao Y (2011) Using transaction prices to re-examine price dispersion in electronic markets. *Inform. Systems Res.* 22(2): 269–288.

Giaglis GM, Klein S, O'Keefe RM (2002) The role of intermediaries in electronic marketplaces: Developing a contingency model. *Inform. Systems J.* 12(3):231–246.

Girvan M, Newman ME (2002) Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* 99(12):7821–7826.

Haynes M, Thompson S (2008) Price, price dispersion and number of sellers at a low entry cost shopbot. *Internat. J. Indust. Organ.* 26(2):459–472.

Hinz O, Skiera B, Barrot C, Becker JU (2011) Seeding strategies for viral marketing: An empirical comparison. *J. Marketing* 75(6):55–71.

Kamada T, Kawai S (1989) An algorithm for drawing general undirected graphs. *Inform. Processing Lett.* 31(1):7–15.

Kamakura WA, Russell GJ (1989) A probabilistic choice model for market segmentation and elasticity structure. *J. Marketing Res.* 26(4):379–390.

Kendall M, Cockel R, Becker J, Hawkins C (1970) Raised serum alkaline phosphatase in rheumatoid disease. An index of liver dysfunction? *Ann. Rheumatic Diseases* 29(5):537–540.

Kim JB, Albuquerque P, Bronnenberg BJ (2010) Online demand under limited consumer search. *Marketing Sci.* 29(6):1001–1023.

Kim JB, Albuquerque P, Bronnenberg BJ (2011) Mapping online consumer search. *J. Marketing Res.* 48(1):13–27.

Lancichinetti A, Fortunato S (2009) Community detection algorithms: A comparative analysis. *Physical Rev. E* 80(5):056117.

Lattin JM, Carrol DJ, Green PE (2003) *Analyzing Multivariate Data* (Duxbury Resource Center, Pacific Grove, CA).

Lee TY, Bradlow ET (2011) Automated marketing research using online customer reviews. *J. Marketing Res.* 48(5):881–894.

Lilien GL, Rangaswamy A (2004) *Marketing Engineering: Computer-Assisted Marketing Analysis and Planning* (DecisionPro, Victoria, BC, Canada).

Marbeau Y (1998) Communication of research results. *The ESOMAR Handbook of Market and Opinion Research* (ESOMAR, Amsterdam), 519–552.

Meunier D, Lambiotte R, Fornito A, Ersche KD, Bullmore ET (2009) Hierarchical modularity in human brain functional networks. *Frontiers Neuroinformatics* 3(37):1–12.

Moe WW (2006) An empirical two-stage choice model with varying decision rules applied to Internet clickstream data. *J. Marketing Res.* 43(4):680–692.

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

Newman ME (2004) Fast algorithm for detecting community structure in networks. *Physical Rev. E* 69(6):066133.

Newman ME, Girvan M (2004) Finding and evaluating community structure in networks. *Physical Rev. E* 69(2):026113.

Newman JW, Lockeman BD (1975) Measuring prepurchase information seeking. *J. Consumer Res.* 2(3):216–222.

Onnela JP, Chakraborti A, Kaski K, Kertész J, Kanto A (2003) Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Rev. E* 68(5):056110.

Ozimec A-M, Natter M, Reutterer T (2010) Geographical information systems-based marketing decisions: Effects of alternative visualizations on decision quality. *J. Marketing* 74(6):94–110.

Palla G, Derényi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435(7043):814–818.

Paulssen M, Bagozzi RP (2006) Goal hierarchies as antecedents of market structure. *Psych. Marketing* 23(8):689–709.

Peter JP, Olson JC (1993) *Consumer Behavior and Marketing Strategy*, 3rd ed. (Irwin, Homewood, IL).

Ramaswamy V, DeSarbo WS (1990) Sculptre: A new methodology for deriving and analyzing hierarchical product-market structures from panel data. *J. Marketing Res.* 27(4):418–427.

Rao VR, Sabavala DJ (1986) *Measurement and Use of Market Response Functions for Allocating Marketing Resources* (Marketing Science Institute, Boston).

Rappuoli R, Aderem A (2011) A 2020 vision for vaccines against HIV, tuberculosis and malaria. *Nature* 473(7348):463–469.

Roberts JH, Lattin JM (1991) Development and testing of a model of consideration set composition. *J. Marketing Res.* 28(4):429–440.

Rotta R, Noack A (2011) Multilevel local search algorithms for modularity clustering. *J. Experiment. Algorithmics* 16:2–3.

Shocker AD, Ben-Akiva M, Boccara B, Nedungadi P (1991) Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing Lett.* 2(3):181–197.

Sipior JC, Ward BT, Mendoza RA (2011) Online privacy concerns associated with cookies, flash cookies, and Web beacons. *J. Internet Commerce* 10(1):1–16.

Smelcer JB, Carmel E (1997) The effectiveness of different representations for managerial problem solving: Comparing tables and maps. *Decision Sci.* 28(2):391–420.

Smith MD, Brynjolfsson E (2001) Consumer decision-making at an Internet shopbot: Brand still matters. *J. Indust. Econom.* 49(4):541–558.

Swait J, Erdem T (2007) Brand effects on choice and choice set formation under uncertainty. *Marketing Sci.* 26(5):679–697.

Tirunillai S, Tellis G (2014) Mining marketing meaning from chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *J. Marketing Res.* 51(4):463–479.

Urban G, Johnson PL, Hauser JR (1984) Testing competitive market structures. *Marketing Sci.* 3(2):83–112.

Urban GL, Hulland JS, Weinberg BD (1993) Premarket forecasting for new consumer durable goods: Modeling categorization, elimination, and consideration phenomena. *J. Marketing* 57(2): 47–63.

van Eck NJ, Waltman L (2007) *VOS: A New Method for Visualizing Similarities Between Objects* (Springer, Berlin).

van Eck NJ, Waltman L, Dekker R, van den Berg J (2010) A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. *J. Amer. Soc. Inform. Sci. Tech.* 61(12):2405–2416.

Vera JF, Macías R, Heiser WJ (2009) A dual latent class unfolding model for two-way two-mode preference rating data. *Comput. Statist. Data Anal.* 53(8):3231–3244.

Yao W-M, Fahmy S (2014) Flow-based partitioning of network testbed experiments. *Comput. Networks* 58:141–157.