

Multimarket Membership Mapping

Daniel M. Ringel

Journal of Marketing Research
 2023, Vol. 60(2) 237-262
 © The Author(s) 2022



Article reuse guidelines:
sagepub.com/journals-permissions
 DOI: 10.1177/00222437221110460
journals.sagepub.com/home/mrj



Abstract

Market structure maps spatially represent competitive relationships among brands and products. In these maps, each product is typically visualized in a single map location. Such visualization, however, does not adequately accommodate multimarket membership products (MMPs), which are products that compete in multiple submarkets that are each characterized by distinct competitors and consumer preferences. The author proposes a novel framework that accommodates MMPs by visualizing products in the local context of each submarket in which they compete. This framework is then used to study market structure in the digital camera market. Competitive relationships are inferred from consumers' online searches using bootstrapped neural product embedding. The research shows that 8% of cameras are MMPs and that 50% of all submarkets are affected by MMPs. The map generated with the proposed framework uncovers mismatches between manufacturer positioning and consumer perception of several cameras. An extensive comparison of the proposed framework with alternative mapping methods reveals that (1) commonly used mapping methods misrepresent MMPs' competitive positions in their maps and (2) ignoring MMPs can lead to distorted visualizations of the underlying market structure. Both problems compromise firms' abilities to accurately assess products' competitive positions and evaluate the effectiveness of their positioning strategies.

Keywords

competitive market structure visualization, overlapping clustering, online search, neural product embedding

Online supplement: <https://doi.org/10.1177/00222437221110460>

Firms rely on the analysis of market structures so that interrelations among products and brands become apparent (Grover and Dillon 1985). A pivotal approach is the derivation of a spatial configuration or map that depicts the continuum of competitive relationships among the products in a given market (DeSarbo, Manrai, and Manrai 1993). A major advantage of market structure maps is that they reveal relationships that would otherwise not be noticeable or straightforward to interpret (Gabel, Guhl, and Klapper 2019; Ringel and Skiera 2016). Gartner refers to “visual-based data discovery” as a “defining feature” of modern business analytics in its Magic Quadrant report (Gartner 2018, p. 2). Information derived from maps translates into concrete managerial consequences, informing marketing-mix and positioning decisions (Lattin and McAlister 1985; Lilien and Rangaswamy 2004; Smelcer and Carmel 1997).

Mapping methods commonly used in marketing, such as the popular multidimensional scaling (MDS) method, assign each product to a single location on a map. Although this property facilitates the embedding of information in maps, it misses important relationships when products or brands compete in multiple submarkets. Research on market segmentation suggests that markets can consist of multiple submarkets and that products can compete in multiple submarkets simultaneously

(Arabie et al. 1981). By positioning each product in a single location on market structure maps, extant mapping methods fail to fully capture the true competitive positions of products that compete in multiple submarkets simultaneously. In this research, I refer to these products as multimarket membership products (MMPs). In what follows, I show that failure to account for these products' multimarket memberships can lead to distorted visual representations of the market structure. This problem can compromise the accurate assessment of products' competitive positions and may mislead firms about the effectiveness of their positioning strategies.

This research aims to overcome the limited capacity of extant mapping methods to accommodate MMPs. Specifically, in this research I (1) develop a flexible framework that identifies and accurately visualizes MMPs in market structure maps, (2) use this framework to investigate multimarket memberships in the digital single-lens reflex (DSLR) camera market, (3) demonstrate the strengths of this framework by a numerical example and a simulation, and (4) compare this framework with alternative mapping methods.

Daniel M. Ringel is Assistant Professor of Marketing, Kenan-Flagler Business School, University of North Carolina at Chapel Hill, USA (email: dmr@unc.edu).

This new framework is referred to as multimarket membership mapping, abbreviated as M4. The M4 framework addresses the shortcomings of existing mapping approaches by creating a single map that captures all submarkets of a given market without making a priori assumptions as to which products to visualize in one or more submarkets. M4 builds on the idea that market segmentation and market structure analysis are complementary (Green, Wind, and Claycamp 1975; Grover and Srinivasan 1987). At its core, M4 connects overlapping clustering, which can detect multimarket memberships of products, with map projection in a novel way. This novel connection enables M4 to (1) identify submarkets, (2) determine which products compete in more than one submarket, and (3) visualize products in the local context of each submarket they compete in on a common market structure map.

The M4 framework is based on the premise that in markets characterized by high levels of product variety, complexity, and versatility, a single product may meet different consumer preferences that were traditionally met by products of different submarkets (Day, Shocker, and Srivastava 1979; Shocker, Bayus, and Kim 2004; Urban, Johnson, and Hauser 1984). A submarket is defined as the subset of a market's products that consumers perceive, on the basis of their preferences, to be more substitutable for one another than for other products in the market. An MMP is thus a product that consumers perceive to be a viable substitute in more than one submarket.

This research demonstrates the value of accurately accommodating MMPs in market structure maps in an empirical study of the DSLR camera market. Using neural product embedding, I infer competitive relationships among DSLR cameras from online search histories (or clickstreams) of over 57,000 consumers at a major price comparison site. Submarkets emerge when multiple consumer segments exist whose search spaces (i.e., consideration sets) are very similar to another, but different to those of other segments.

The M4 framework reveals that among the 111 DSLR cameras in this study, 8% compete in the context of more than one of ten identified submarkets. Half of all submarkets are affected by the identified MMPs. A deeper analysis of the underlying clickstream data confirms that a large majority of consumers who search for an MMP do so within the context of either one or the other of its submarkets, but not across submarkets. Importantly, the map produced by M4 indicates where manufacturers' intended camera positions (as stated in their product descriptions, for example, "semi-professional camera") are inconsistent with consumers' perceptions. The comparison of M4's market structure map of the DSLR camera market to maps of alternative methods shows that M4 outperforms them in terms of face validity and the extent to which it accurately represents the underlying competitive relationships in its map. An extensive simulation study supports these findings.

The rest of this article is organized as follows: The next section provides an overview of related literature. Then, I

outline the basic idea of the M4 framework, followed by a detailed description and a numerical example. I explain how competitive relationships can be inferred from online search data using neural product embedding and then analyze multimarket memberships in the DSLR camera market. After testing the validity and robustness of the findings, I compare the M4 map with maps of alternative methods empirically and by simulation. Finally, I highlight the contributions of this research and close with a discussion of implications, limitations, and future research directions.

Related Literature on Mapping and Clustering

A market structure map is a two-dimensional plane on which product relationships are visualized.¹ On such maps, often produced using MDS or alternative map projection methods (for a review, see France and Ghose 2018), competitive relationships among products are represented through spatial proximity. Products identified as having strong competitive relationships are located close to one another, whereas those with weak relationships are distant. The use of maps to visualize competitive relationships is deeply ingrained in the marketing literature (e.g., Andrews and Manrai 1999; DeSarbo, Grewal, and Scott 2008; Elrod 1988; Green 1975).

Marketers regularly use MDS and its many extensions to study consumer perceptions and investigate market structure. Since the introduction of MDS, a steady stream of mapping methods evolved to account for more complex data sets and more advanced statistical requirements (France and Ghose 2018). Many examples that leverage map-based data visualization to investigate market structures exist. These include mapping of consumer online search (Kim, Albuquerque, and Bronnenberg 2011), market structure analysis from discussion forums (Netzer et al. 2012), visualization of asymmetric competition in large markets (Ringel and Skiera 2016), and visualization of submarkets from purchase data (France and Ghose 2016). Further examples involve the use of statistical perceptual maps for inference (Iacobucci, Grisaffe, and DeSarbo 2017), visualization of brand relationships from social tags (Nam, Joshi, and Kannan 2017), visualization of large retailer assortments from shopping basket data (Gabel, Guhl, and Klapper 2019), and visualization of market structure on the basis of social media engagement (Yang, Zhang, and Kannan 2021).

Historically, cluster analysis and market structure visualization were largely addressed in separate streams of literature that developed independently of one another. Whereas cluster analysis provides set information (i.e., submarket memberships),

¹ Higher-dimensional representations of product relationships are also possible. However, previous research suggests that adding even a third dimension should be avoided wherever possible because it complicates viewing and interpretation (e.g., Marbeau 1998).

market structure maps provide a continuous perspective on products' relative positions. Thus, maps qualify competitive relationships further than clustering solutions do because they reveal the extent to which products compete with one another (France and Ghose 2016; Shocker, Stewart, and Zahorik 1990).

Several studies highlight the value of combining mapping with clustering. For instance, Netzer et al. (2012) and France and Ghose (2016) use cluster analysis to identify products' submarket memberships on market structure maps. Likewise, Gabel, Guhl, and Klapper (2019) cluster the products in their assortment map to better visualize store categories from consumers' perspectives. Although the posterior consideration of submarket memberships in these studies does not inform the actual mapping process, it facilitates the interpretation of the generated maps.

Closer to this work are studies that integrate clustering and mapping into a single method, thereby deriving additional benefits. For instance, the clusterwise bilinear spatial method proposed by DeSarbo, Grewal, and Scott (2008) performs segmentation and positioning simultaneously. Ringel and Skiera (2016) integrated clustering and mapping in a joint method called DRMABS to extend market structure visualization to large markets that comprise dozens of submarkets and hundreds of products. However, neither method identifies MMPs and embeds them accurately in the submarkets that they compete in on their maps.

Whereas extant mapping methods assume that products have a single position in a market, research on market segmentation and cluster analysis shows that this assumption does not always hold. Many products are, in fact, MMPs (Arabie et al. 1981; Chaturvedi et al. 1997; Kannan and Sanchez 1994; Wedel and Steenkamp 1991; Zahorik 1994). Therefore, an approach that acknowledges this aspect of products' competitive positions is necessary.

Because extant mapping methods can only assign products to single map locations, analysts have had two options for visualizing market structures in the presence of MMPs. The first option involves visualizing MMPs somewhere between their respective submarkets. Yet, such an approach can disguise the local competitive roles of MMPs and may further distort the global representation of the underlying market structure—potentially suggesting competitive relationships among products when no such relationships exist.

The second option is to create an individual map for each submarket. However, the resulting series of independent maps are cumbersome to study and fall short of uncovering markets' global structures, how submarkets relate to each other, and where products compete across submarkets' boundaries (Ringel and Skiera 2016). Further, to generate a series of independent maps, the analyst would need to first define submarkets and then make assumptions as to which submarkets any given MMP belongs to from consumers' perspectives. The prerequisite for such approach is at least some familiarity on the part of the analyst with all competitors in each submarket—a daunting task in larger markets (Gabel, Guhl, and Klapper 2019).

In contrast, the M4 framework explicitly acknowledges the possibility of products competing in the local contexts of more than one submarket. Using the M4 framework, analysts can generate a single market structure map that comprises all submarkets of a given market and visualizes MMPs in the local competitive contexts of each of their submarkets.

In terms of data, this research relates most closely to the work of Ringel and Skiera (2016) and Kim, Albuquerque, and Bronnenberg (2011). Both studies also use consumer search data to infer competitive relationships among products. However, neither study considers MMPs in its market structure maps, nor do these studies engage in a deeper analysis of consumers' search patterns to investigate products' map positions. Furthermore, whereas previous studies provide theoretical support for using consumer search data to identify competitive relationships from co-consideration (e.g., Kim, Albuquerque, and Bronnenberg 2011; Moe 2006; Ringel and Skiera 2016), this research is the first to validate the notion externally by surveying consumers (see the "Map Validity" subsection in the "Empirical Study of the DSLR Camera Market" section).

Finally, although I use price comparison site data as Ringel and Skiera (2016) do, I infer competitive relationships from these data differently using neural product embedding. In this regard, this research is more closely related to that of Gabel, Guhl, and Klapper (2019), who implement a variation of word2vec (Mikolov et al. 2013), a shallow neural network for natural language processing, to infer relationships among consumable products of a grocers' assortment from scanner data. In contrast to the approach of Gabel, Guhl, and Klapper, I use a bootstrapping procedure to ensure solution robustness, and I explicitly consider product sequence in search (which scanner data do not capture). Importantly, while their analysis reveals product categories from consumers' perspectives, it does not consider the possibility that a consumable can be a substitute (or complement) in more than one category. Table 1 provides an overview of studies that are most closely related to this research and highlights key differences and contributions.

Basic Idea of the M4 Framework

The M4 framework is the first mapping approach that explicitly accommodates MMPs in market structure maps. Defining the market under analysis is a prerequisite for visualizing competitive market structure. Markets can be characterized by various parameters, such as geography, product category, regulation, politics, social structure, firms, customers, technology, North American Industry Classification System codes, or combinations thereof. For the purpose of this research and without loss of generality, I follow France and Ghose's (2016) characterization. Their research views market competition at the product category level, where each product is typically a member of a submarket.

The definition of a submarket used in this research—the subset of a market's products that consumers perceive to be more substitutable to one another than to other products in the market on the basis of their preferences—accommodates

Table 1. Overview of Related Marketing Research.

Study	Data	Clustering	Mapping	MMPs	Large Markets ^a	Competitive Asymmetry	Consumer Heterogeneity	Automatic Submarket Detection ^b	Map Not Subject to Lumping and Circular Bending ^c	Safeguards to Ensure Map Robustness
Wedel and Steenkamp (1991)	Consumer interviews	Fuzzy clusterwise regression	—	Yes	—	—	—	Yes	—	—
Kannan and Sanchez (1994)	Scanner data	Subset selection	—	Yes	—	Yes	—	—	—	—
Zahorik (1994)	Scanner data	Nonhierarchical brand switching	—	Yes	—	—	Explicitly modeled	—	—	—
Chaturvedi et al. (1997)	Conjoint analysis	Overlapping k-centroids	—	—	—	—	—	—	—	—
DeSarbo, Grewal, and Scott (2008)	Product attribute ratings	Clusterwise bilinear MDS	MDS	—	—	—	Explicitly modeled	Yes	—	—
Kim, Albuquerque, and Bronnenberg (2011)	Also viewed product lists	—	Asymmetric MDS	—	—	Yes	—	—	—	—
Netzer et al. (2012)	Posts in forums	Community detection	Force-directed graph or MDS	—	Yes	—	—	Yes	—	—
France and Ghose (2016)	Scanner data; restaurant reviews	SUBMARIT	MDS	—	Yes	—	—	—	—	—
Ringel and Skiera (2016)	Online search	Community detection	DRMABS	—	Yes	Yes	Implicitly captured ^d	Yes	Yes	—
Iacobucci, Grisaffe, and DeSarbo (2017)	Consumers' similarity and familiarity judgements	—	INDSCAL with confidence regions	—	—	—	Explicitly modeled ^a	—	—	—
Nam, Joshi, and Kannan (2017)	Social tags (keywords)	Parameterized finite Gaussian mixture model	Force-directed graph	—	Yes	—	—	—	—	—
Gabel, Guhl, and Klapper (2019)	Scanner data	k-Means	t-SNE	—	Yes	—	—	—	Yes	—
Yang, Zhang, and Kannan (2021)	Likes and posts on brands' social media pages	k-Means	t-SNE	—	Yes	—	—	—	Yes	—
This study	Online search	SpeakEasy	t-SNE	Yes	Yes	—	Implicitly captured ^d	Yes	Yes	Yes

^aMarkets with 100+ products.^bAutomatically detect number of submarkets without additional heuristics.^cKnown limitations of MDS and force-directed graphs in larger markets and with sparse data.^dEmbedded in consumers' individual searches (see the "Accommodating Consumer Heterogeneity" subsection in the "Empirical Study of the DSLR Camera Market" section for additional details).

the idea that markets, even entire economies, can be understood as interlaced networks of relationships (Thorelli 1986). In this sense, a submarket's products compete more strongly with one another than with products of other submarkets (France and Ghose 2016). By accounting for the possibility that products may also maintain competitive relationships to products outside of their submarkets, market structure can be visualized as a continuum of product positions within and across submarkets on a map.

Therein, an MMP is a product that consumers perceive to be a viable substitute in more than one submarket. An example of an MMP is the luxury SUV that Van Heerde, Srinivasan, and Dekimpe (2010) study. Their analysis reveals that the luxury SUV competes in two submarkets: luxury limousines and SUVs. In each submarket, the luxury SUV competes on the specific consumer preferences pertaining to that submarket and faces a different set of competitors. Because common market structure maps assign products to single map locations, they cannot visualize MMPs such as the luxury SUV in the local competitive context of each of their submarkets on the same map.

The basic idea of the M4 framework for accurately accommodating MMPs in market structure maps is to connect overlapping clustering with map projection. These two methods may, however, seem inherently incompatible: overlapping clustering receives unique objects as input and assigns them to multiple clusters, which effectively removes objects' uniqueness. Map projection methods, in contrast, assume that objects are unique.

To reconcile overlapping clustering with map projection, I propose to represent each MMP as a set of unique instances. Each instance corresponds to one of the submarkets to which the MMP is assigned by the overlapping clustering method. To capture the MMP's competitive position within the specific local contexts of each of the submarkets it competes in, all competitive relationships are removed from each of its instances that do not pertain to the submarket that the instance is associated with. In other words, MMPs' competitive relationships are distributed to other products in the market across their respective instances such that each instance only captures an MMP's local competitive position. Because each instance of an MMP is now a unique object in the map projection process, the MMP can appear in multiple submarkets, that is, locations on the same map.

Four Phases of the M4 Framework

This basic idea of this research is implemented in the four-phase M4 framework. The four consecutive phases, shown in Figure 1, are (1) input preparation, (2) identification of multimarket memberships, (3) transformation of multimarket memberships, and (4) map projection. The output of each phase serves as direct input to the next phase. I explain each phase of M4 in detail in the following four subsections.

Phase 1: Input Preparation

The M4 framework operates on a relationship matrix that captures the presence and magnitude of pairwise relationships

among products in higher-dimensional space. For market structure analysis, relationship matrix \mathbf{S} captures the extent to which the products of a market compete with one another. Assume that J is an index set of all products in a market under consideration, $J = \{1, \dots, |J|\}$. The input to M4 is a $|J| \times |J|$ relationship matrix \mathbf{S} , in which each cell S_{ij} , $i, j \in J$, captures the competitive relationship between product i and product j . Specifically, S_{ij} takes a value from 0 to 1, such that higher values indicate heavier competition than lower values. For $i = j$ it is assumed that $S_{ij} = 0$, since products do not compete against themselves. M4 does not require symmetric relationships; that is, relationship matrix \mathbf{S} does not necessarily abide by metric axioms such as triangle inequality.

To construct relationship matrix \mathbf{S} , it is necessary to infer competitive relationships among product pairs (i, j) on the basis of relevant data sources that are available to the analyst. Previous research presents numerous approaches to infer competitive relationships from various data sources. Some examples include text mining of user-generated content (Lee and Bradlow 2011; Netzer et al. 2012), co-occurrence in consumer search (Kim, Albuquerque, and Bronnenberg 2011; Ringel and Skiera 2016), brand switching in purchase panel data (France and Ghose 2016), and co-occurrence in consumers' shopping baskets (Gabel, Guhl, and Klapper 2019). Notably, the features (or attributes) that drive the presence and magnitude of products' relationships can be latent; that is, they need not be directly observed for the purpose of market structure analysis (e.g., Gabel, Guhl, and Klapper 2019). In the empirical study of the DSLR camera market, relationship matrix \mathbf{S} is constructed from consumers' clickstreams using neural product embedding.

Phase 2: Identification of Multimarket Memberships

The objective of the second phase of M4 is to identify products that simultaneously compete in multiple different submarkets, that is, against different sets of local competitors. The second phase involves (1) identifying submarkets of competing products from relationship matrix \mathbf{S} and (2) determining which products compete in several of these submarkets.

I use an overlapping clustering method to identify overlapping submarkets of products, that is, markets where products may be members of more than one submarket. In contrast to standard clustering methods that produce disjoint solutions in which each product can only be assigned to one submarket, overlapping clustering methods can produce nondisjoint solutions; they can identify which products belong to more than one submarket and assign them accordingly.²

For the purpose of this study, a clustering method called SpeakEasy is integrated into the M4 framework. SpeakEasy outperforms alternative clustering methods (Gaiteri et al. 2015), and it is particularly well suited for the M4 framework

² See Web Appendix A for an overview and the history of overlapping clustering in marketing.

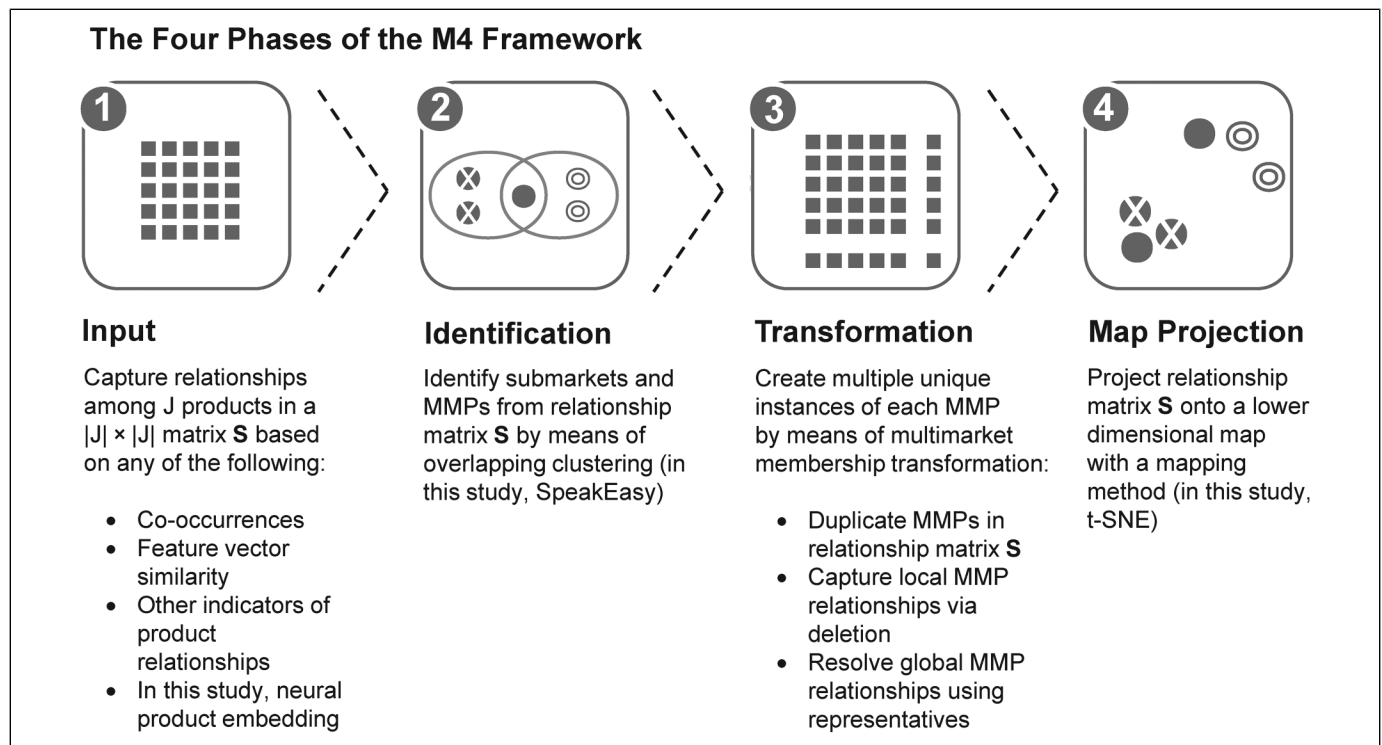


Figure 1. The Four Phases of the M4 Framework.

because it (1) operates directly on a relationship matrix, (2) identifies the number of clusters automatically, (3) identifies MMPs, (4) produces robust solutions, and (5) is computationally inexpensive.

SpeakEasy is based on the principle of label propagation, where objects join clusters through an exchange of “labels” between related objects. In label propagation, objects are initially randomly assigned to a network, where each object holds a single label; with each iteration of the algorithm, the object’s label is updated to match the label carried by the majority of the object’s neighbors in the network. Clusters are identified when the algorithm converges.

A novelty of SpeakEasy’s label propagation process is that it combines a bottom-up approach to clustering (using information regarding network neighbors) with a top-down approach (using information from the whole set of objects). Specifically, SpeakEasy (1) updates each object’s labels on the basis of its neighbors’ labels and (2) subtracts the expected frequency of these labels based on their prevalence in the complete set of objects. Such a dual approach facilitates accurate cluster detection because the top-down information is used to ensure that the bottom-up label propagation process identifies clusters that accurately represent the global structure (De Meo et al. 2014). SpeakEasy’s label propagation process thus serves the objective of capturing both the global structure of the market and local submarket structures in the maps.

Label propagation is a greedy algorithm, which makes SpeakEasy extremely fast but also leads to different clustering solutions across runs. To ensure solution robustness,

SpeakEasy rapidly executes numerous clustering runs and applies consensus clustering (Lancichinetti and Fortunato 2012) to find the clustering solution that is most representative of all identified solutions. By evaluating the oscillation of objects between clusters across all identified solutions relative to the representative solution, SpeakEasy can identify cluster overlaps. Specifically, SpeakEasy tracks objects’ pairwise cluster assignments across all solutions in a weighted co-occurrence matrix. Using an adaptive threshold (t) set as a function of the maximum allowable number of clusters that an object can be assigned to (r) in conjunction with the weighted co-occurrence matrix, SpeakEasy identifies objects that belong to multiple clusters. The underlying intuition is that as the number of clusters an object can be associated with increases, its weight within each cluster will decrease. Thus, the maximum number of clusters permissible, as expressed by the threshold $t = 1/r$, is a constraint that provides some control over the degree of cluster overlap. Because of the possibility that products compete in more than two submarkets, r is set to 3, such that more than two submarket assignments are possible.

Arguably, solution robustness is desirable for market structure analysis. However, so is solution quality, which SpeakEasy does not explicitly consider. Because label propagation is greedy in nature, it can easily get trapped in local optima. In consequence, poor clustering solutions end up in SpeakEasy’s consensus clustering phase, where they affect the choice of the representative clustering solution unfavorably. In other words, the chosen representative solution will share

properties with poor solutions. To overcome this problem, I introduce a solution quality criterion called modularity to SpeakEasy's consensus clustering phase.³ Modularity is the most widely used measure of solution quality in the area of community detection (Sobolevsky et al. 2014). It compares the weighted number of relationships inside a cluster with the expected value for a randomized network of the same size and degree sequence. Clustering solutions with high modularity are considered good because they are different from a random assignment of objects to clusters. Thus, modularity is a quantitative measure of the quality of the division of a network into clusters or, in the case of market analysis, the division of a market into submarkets. In this modification of SpeakEasy, only clustering solutions with higher modularities are considered in the consensus clustering phase. Median modularity is used as a heuristic to divide the better from the worse clustering solutions of all clustering runs. This modified SpeakEasy method identifies the representative clustering solution only from a pool of higher-quality solutions (i.e., with modularities above the median). Thus, although the final clustering solution retains its robustness, its quality is higher because it does not need to represent poorer solutions that arise from the greedy nature of label propagation.

Phase 3: Transformation of Multimarket Memberships

In general, map projection methods receive relationship matrices (such as the original input matrix S) as input and translate each relationship between two objects into a Euclidean distance on a map. Accordingly, the objective of the third phase of M4 is to transform relationship matrix S , coupled with the output of the multimarket membership identification phase—that is, assignments of each product into (potentially) multiple submarkets—into a single matrix that captures the products' multimarket memberships, such that it can be input into a map projection method (in Phase 4).

To this end, I propose a new process called multimarket membership transformation (MMT). Broadly, this process entails modifying relationship matrix S so that each MMP identified in the previous phase is represented by multiple unique instances in the matrix. Each instance represents the local competitive relationships between that MMP and the other products within one of its specific submarkets. In what follows, I provide a general overview of the MMT process and illustrate its intuition in Figure 2. The input in Figure 2 is relationship matrix S from Phase 1 and the overlapping clustering solution from Phase 2 of the M4 framework. I describe the MMT process formally in the Appendix and provide a detailed step-by-step numerical example in Web Appendix B.

Step 1: Create multiple instances for each MMP in the matrix. The first step in MMT is to duplicate the row and column of each MMP in the relationship matrix S (i.e., to create a new instance for each MMP) until the number of instances of the MMP is equivalent to the number of submarkets that the MMP competes in (as identified in the previous phase of M4). After this step is completed, each MMP's corresponding instances are identical. In other words, they do not yet capture the MMP's specific local competitive relationships to other products in particular submarkets. Thus, it is necessary to ensure that (1) the competitive relationships represented in each instance are context specific (i.e., pertain to a specific submarket that the MMP competes in locally) and (2) competitive relationships are not artificially inflated (i.e., are not duplicated across submarkets). To this end, MMT divides the competitive relationships of a given MMP across its instances in the second step.

Step 2: Capture local relationships for each MMP instance. Consider a focal MMP, denoted MMP j , and a focal instance corresponding to the MMP. Recall that this instance is ultimately supposed to represent the local competitive relationships between MMP j and the other products pertaining to a specific submarket in which MMP j competes (for convenience, I refer to that submarket as submarket A). Thus, all competitive relationships that pertain to MMP j 's other submarkets—that is, in all of MMP j 's submarkets except submarket A—are removed from the instance.


This step is carried out for each instance (pertaining to a specific submarket) of MMP j and for all instances of other MMPs in the analysis. By the end of this process, the instances corresponding to each MMP will be different from one another, with each instance capturing the local competitive position of its MMP within the context of a specific submarket.

Step 3: Resolve global relationships. This step returns to the instance of MMP j that corresponds to submarket A. The previous step eliminated from that instance all local relationships to products that compete with MMP j in its other submarkets (i.e., that do not compete in submarket A). However, the possibility remains that MMP j also weakly competes with products that are not part of any of its submarkets.

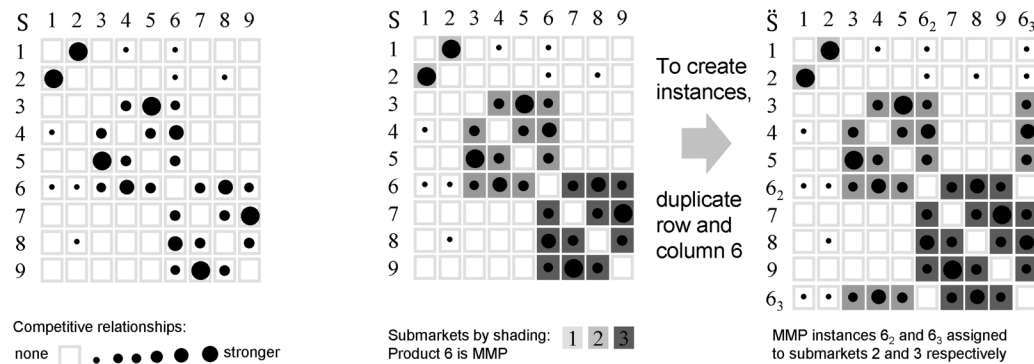
These global competitive relationships are not accounted for by the previous step and are thus still duplicated across MMP j 's instances. To resolve the duplication and thus artificial inflation of global competitive relationships, each of them must be assigned uniquely to MMP j 's instances. Substantively, global competitive relationships should be assigned to the instance of an MMP whose competitive context they pertain to (that is, submarket and thus benefits sought by consumers of that submarket). For the aforementioned luxury SUV, a global competitive relationship to a vehicle that is neither a luxury limousine nor an SUV should thus be assigned to the instance of the submarket whose vehicles are most similar to it. Accordingly, on the basis of the premise that each global relationship is assigned to a single instance of MMP j and is eliminated

³ Modularity (Q) = $\frac{1}{2m} \sum_{ij} \left(S_{ij} - \frac{k_i k_j}{2m} \right) \delta(C_i, C_j)$, where S_{ij} represents the relationship weight (or strength) between products i and j , k_i and k_j are the sums of the weights of the relationships attached to products i and j respectively, $2m$ is the sum of all relationship weights in the network, C is a community (submarket), and δ is the Kronecker delta function.

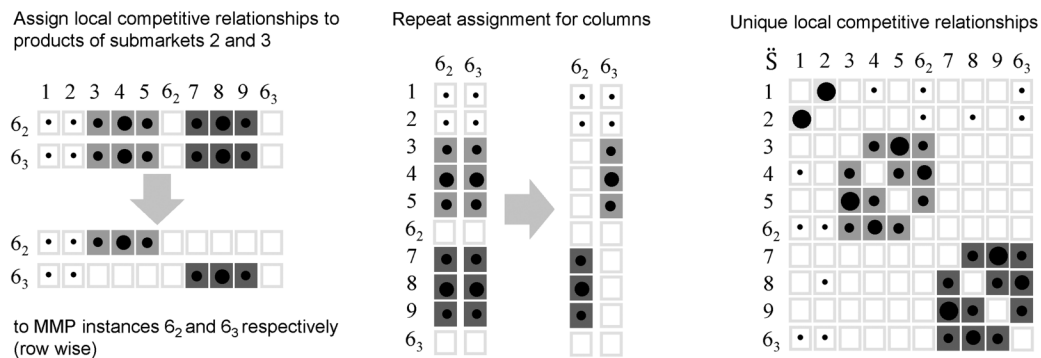
Numerical Example: Multimarket Membership Transformation (MMT)

The competitive relationships of 9 fictitious products (1–9) are captured in a relationship matrix **S**. Each (labeled) row and column represents one of the products. Competitive relationship strength between pairs of products is indicated by bubble size at the intersection of their respective rows and columns. For instance, the competitive relationship of products 1 and 2 is , as captured by matrix cell $S_{1,2}$ (and matrix cell $S_{2,1}$).

Step 1: Create multiple instances for each MMP in relationship matrix **S**



Step 2: Capture local relationships for each MMP instance



Step 3: Resolve global relationships

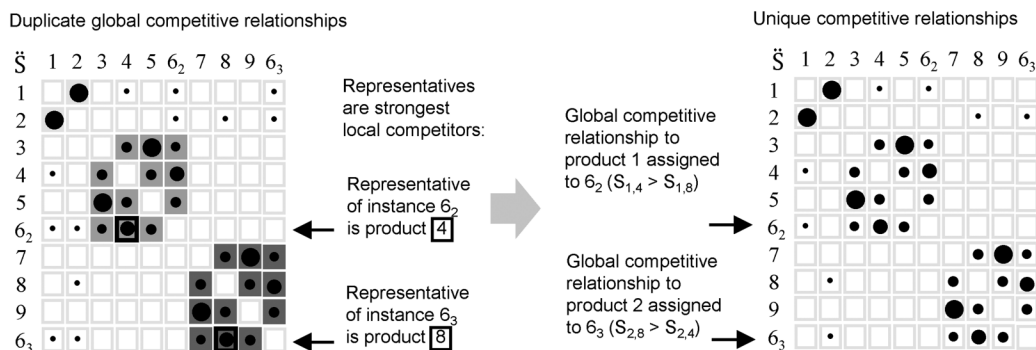


Figure 2. Numerical Example of MMT.

Notes: The input for the MMT process is the relationship matrix **S** from Phase 1 of the M4 framework and the overlapping clustering solution from Phase 2 of the M4 framework.

from the other instances of MMP j , I adopt the following approach: Assume that MMP j has a global relationship with product g . For each instance of MMP j , first the product within the instance's submarket that is most similar to MMP j (that is, with the strongest competitive relationship to MMP j) is identified. This product is referred to as the "representative" of that particular instance of MMP j . The next step is to go back to the original relationship matrix S and identify the representative, among all the representatives of MMP j 's various instances, with which product g has the strongest competitive relationship. This will be the instance to which the MMP's global competitive relationship with product g is assigned. The MMT process also accommodates cases where instances have multiple representatives (i.e., multiple competitors with equally strong competitive relationships to the instance); see the formal description of the MMT process in the Appendix. At the end of the MMT process, both local and global competitive relationships of each MMP are assigned to the submarket-specific instances they pertain to.

Phase 4: Map Projection

In its final phase, M4 outputs a two-dimensional market structure map based on modified relationship matrix S using a map projection method. For this research, a map projection method called t-SNE (van der Maaten and Hinton 2008) is used. To project relationships among objects onto a two-dimensional map, t-SNE uses gradient descent to minimize the Kullback–Leibler (KL) divergence⁴ between object relationships of matrix S and objects' map proximities. Although the KL divergence preserves local structures because of the large penalty it imposes on moving strongly related objects apart, t-SNE's use of a Student's t distribution in its optimization process considers global structure by allowing weakly related objects to be far apart. Because t-SNE is subject to local optima, many iterations and multiple random initializations are run, and the solution with the best fit in terms of t-SNE's cost function is selected. Note that t-SNE is used in this phase because it performs well in the empirical setting (see the results in the "Empirical Study of the DSLR Camera Market" section). The M4 framework can, however, easily accommodate alternative mapping methods, demonstrated in Web Appendix C.

Numerical Example

This section demonstrates the M4 framework by means of a numerical example (as illustrated in Figure 2). The example is sufficiently small in terms of the number of products to allow for a straightforward comparison between actual relationships (represented in matrix S) and visualized relationships (on a

map). The mapping output of M4 is further compared with that of alternative mapping methods to provide a better impression of how the M4 framework contributes to market structure analysis. The objectives are to determine (1) whether M4 visualizes an MMP's competitive position in a market more accurately than alternative mapping methods do and (2) whether M4's explicit consideration of MMPs comes at the cost of distorting competitive relationships among other products in its map.

For the numerical example, the competitive relationships among nine products are simulated as shown in relationship matrix S of Figure 3. The rows and columns of relationship matrix S represent unique products $j \in \{1, \dots, 9\}$. Matrix S_{ij} represents the competitive relationship between products i and j . SpeakEasy identifies three submarkets from relationship matrix S in Figure 3, with Product 6 competing in two of them. The identified submarkets are easily discernable on M4's map in Figure 3. Product 6, an MMP, is visualized in the context of the two submarkets it competes in (that is, Submarkets 2 and 3). Local competitive relationships between pairs of products are well recovered on the map, where products that have strong competitive relationships in matrix S (for instance, Products 1 and 2) appear close to each other, and products that have no competitive relationships at all (for example, Products 4 and 8) appear far apart.

Subsequently, four alternative mapping methods are applied to relationship matrix S to determine to what extent their map projections also correctly recover both local and global competitive relationship structures in the presence of an MMP (see Figure 4). Readers can interactively view all maps of the numerical example at www.mapxp.app/SIM.

Classic MDS is used as a baseline in this comparison. The result shows that classic MDS suffers from circular bending and lumping of products, which disguises the role Product 6 plays in its two submarkets. Moreover, local submarket structures are difficult to discern. To better capture local submarket structures, the comparison includes VOS (Van Eck et al. 2010), a special case of weighted MDS that is related to Sammon mapping (Sammon 1969). VOS weighs relationships by themselves in its optimization process so that weak relationships are considered proportionally less than strong relationships in its map projection. Although VOS does not lump the products corresponding to Submarkets 2 and 3 together, it still isolates Product 6 between its two submarkets. The VOS map also suggests that Products 1 and 2 have a much weaker competitive relationship to each other than is the case.

The more recently introduced DRMABS method (Ringel and Skiera 2016) produces a misleading solution that visualizes Product 6 in only one of its submarkets, thereby suggesting that the MMP does not compete with the products of its other submarket. DRMABS does, however, visualize Products 1 and 2 closer together than MDS and VOS do, which more accurately represents their underlying competitive relationship in matrix S . The fourth mapping method in the comparison is t-SNE, which is also used in M4's final phase in this research. When applied to matrix S (rather than the modification of S derived through M4), t-SNE locates Product 6 between the two submarkets it

⁴ The t-SNE cost function corresponds to $C = \sum_i \text{KL}(P_i || Q_i) = \sum_i \sum_j p_{j|i} \log p_{j|i} / q_{j|i}$ where i and j are objects (e.g., products), p corresponds to objects' (probabilistic) similarities, and q corresponds to objects' (probabilistic) proximities on the map.

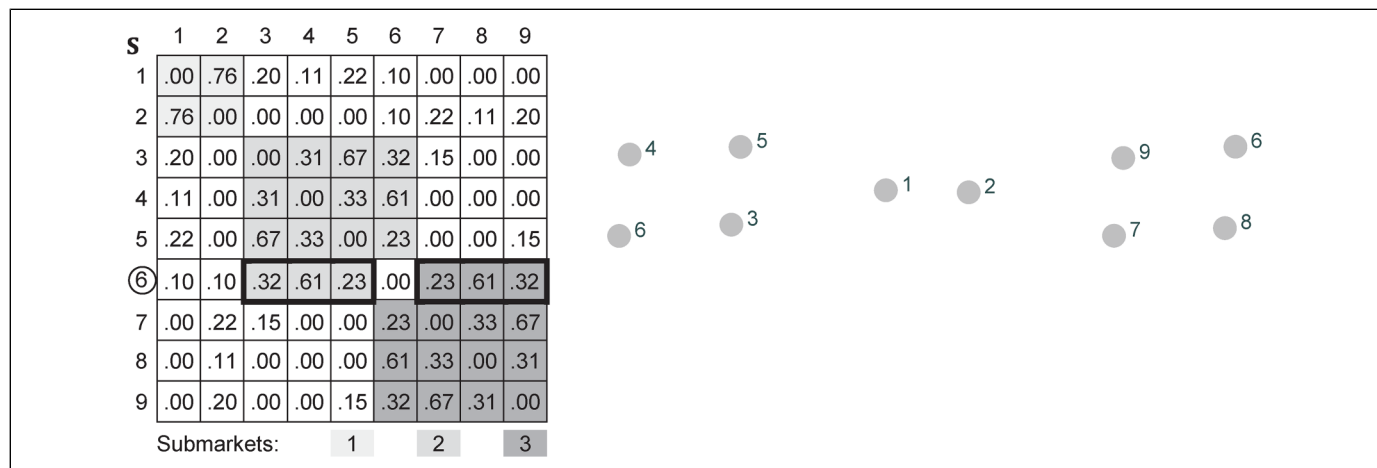


Figure 3. M4 Map Based on Numerical Example.

Notes: Bubbles represent products. Labels identify products of relationship matrix **S**. Axes are on the same scale.

competes in. More severely, the t-SNE map suggests that Products 4 and 8 are competitors when they compete in the context of two different submarkets and have no competitive relationship to each other at all ($S_{4,8} = .00$). Further, the t-SNE map erroneously suggests that Products 3 and 1 are closer competitors than Products 5 and 1 are when the opposite is true ($S_{3,1} = .20$ vs. $S_{5,1} = .22$). Like DRMABS, t-SNE does, however, visualize local competitive relationships among products more accurately than MDS and VOS do.

Common to all four alternative mapping methods is that they push Submarket 1 with Products 1 and 2 off to the side on their maps. M4, in contrast, visualizes Submarket 1 between the other two submarkets on its map (see Figure 3). Because the products of Submarket 1 have equal competitive relationships with products of the two other submarkets, the location of Submarket 1 between the other two submarkets in M4's map seems intuitively correct. Nonetheless, I test whether the global market structure suggested by M4 naturally exists or is artificially imposed by M4's MMT process. To do so, I eliminate Product 6's multimarket membership by removing its competitive relationships to the products of Submarket 2 in a new relationship matrix S^* (see Figure 5). M4 is then applied to this modified relationship matrix S^* .

M4 again identifies three submarkets but no MMP, which is the expected outcome. The previously identified global market structure in which Submarket 1 was located between the other two submarkets persists in Figure 5, confirming that M4's MMT did not impose any artificial structure. Note that with no MMPs present, M4 and t-SNE produce identical maps. In contrast to t-SNE's map in Figure 4, previously misrepresented competitive relationships are now accurately visualized in Figure 5: Products 4 and 8 are far apart, and Products 5 and 1 are closer to each other than Products 3 and 1 are.

The provided numerical example illustrates M4's ability to visualize both local and global market structures more accurately under the presence of an MMP than alternative mapping methods can. More broadly, the numerical example

reveals that the visualization of MMPs in unique map locations can misrepresent their competitive situation, and it can lead to distorted representations of local and global relationship structures among other products.

The numerical example is, however, only an abstraction of real-world markets, which would typically comprise many more products, more submarkets, and perhaps more MMPs, and are likely to have additional relatively weak (or noisy) relationships among products. Together, these factors introduce additional complexity to the relationship structures among products, which most likely hinders obvious scenarios such as the one shown in the provided numerical example. To better understand whether the map distortions suggested by the numerical example might exist in natural empirical settings, additional analysis in the context of the DSLR camera market is presented in Web Appendix D. The alternative methods' maps of the DSLR camera market are subject to distortions similar to those in this numerical example.

Finally, the usefulness of the proposed M4 framework is contingent on the existence of submarkets. This raises the question of how strongly separated the relationship structures (in matrix **S**) should be to qualify as submarkets. Answering this question is not trivial, because the amount of separation that is required depends on the objectives of the particular analysis and what the analyst deems acceptable in the context of a particular market under study. To provide an orientation, the separation between the known submarkets in this numerical example (i.e., in matrix **S**) is reduced with increasing noise levels until they are no longer recovered correctly. After each increase in noise level, the data are reclustered with SpeakEasy, and the modularity of the clustering solution is calculated. The original submarkets in the numerical example are no longer consistently recovered when the modularity drops below .20. Here, the assignment of products to submarkets becomes arbitrary, which (1) compromises the validity of the discovered market structure and (2) thwarts the ability to faithfully identify MMPs from oscillating submarket memberships

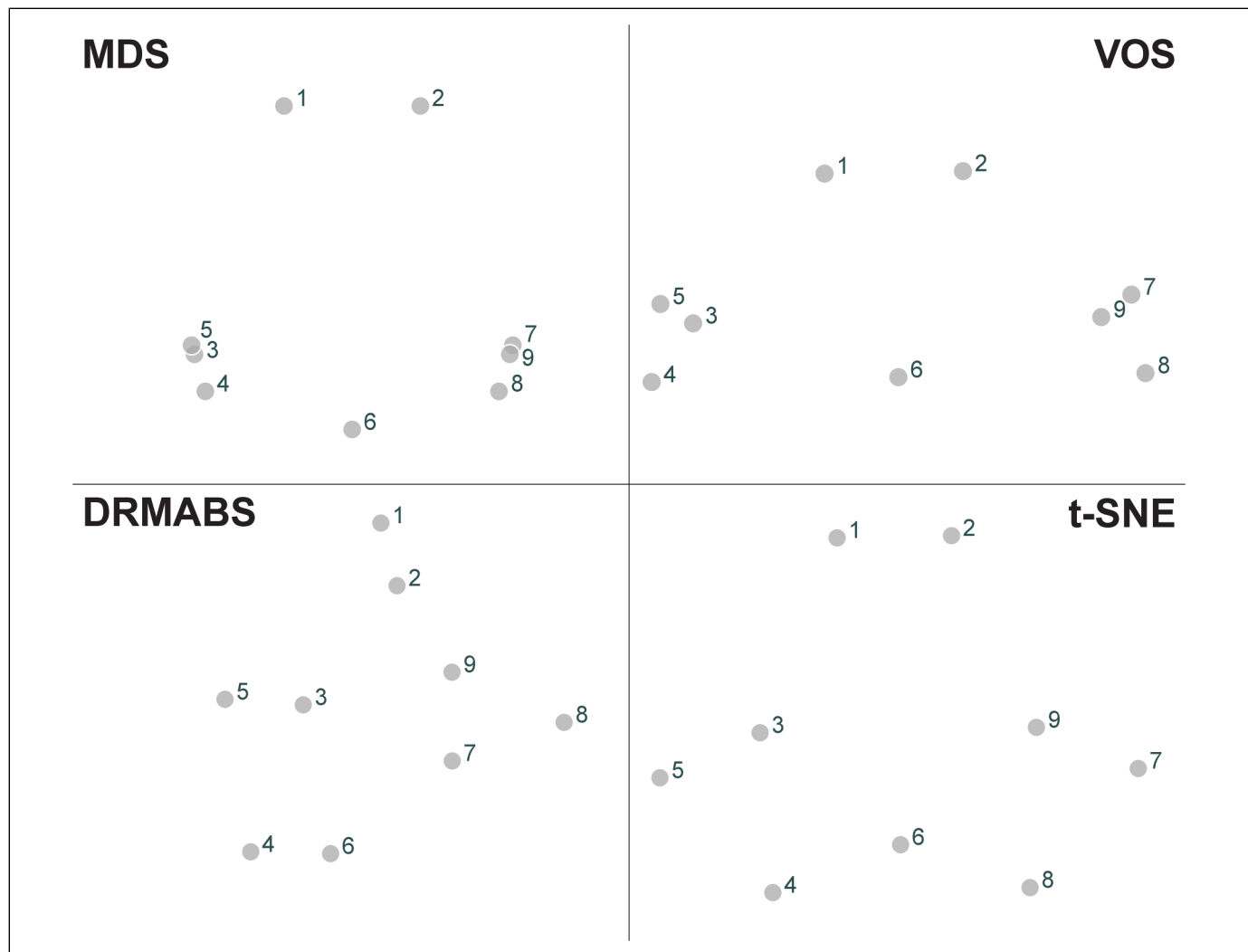


Figure 4. Alternative Methods' Maps Based on Numerical Example.
Notes: Bubbles represent products. Axes are on the same scale.

across multiple clustering runs. Thus, the M4 framework should only be used at modularities well above .20.

Empirical Study of the DSLR Camera Market

An investigation of the market structure in the DSLR camera market demonstrates the value of using M4 over alternative mapping methods in an empirical setting. DSLR cameras are the highest-quality cameras available to consumers. They are used by a broad range of photographers, from aspiring amateurs to professionals. In many cases, camera accessories such as interchangeable lenses are compatible across different models, provided that they are of the same camera system, which is brand specific; consequently, when purchasing a new camera, a consumer who owns such items may be locked into a specific brand.

In what follows, I (1) identify competitive relationships among cameras from consumers' online searches using neural product embedding, (2) use M4 to visualize the underlying market structure, (3) investigate identified multimarket

memberships, (4) test the validity of M4's map, (5) test this approach for robustness against variations in the underlying search data, and (6) compare M4's map with maps produced by alternative mapping methods.

Data

This empirical study follows the work of Ringel and Skiera (2016), who show that consumers' searches on price comparison sites are a strong proxy for market outcome and that competitive relationships among products can be inferred from clickstreams of consumers searching for products online. Clickstream data of 57,144 consumers searching for DSLR cameras at a price comparison website were collected during the first quarter of 2016. These consumers searched for 111 DSLR cameras produced by nine brands. The website's server log was analyzed because it contains the sequence of product views for each unique visitor across multiple sessions.

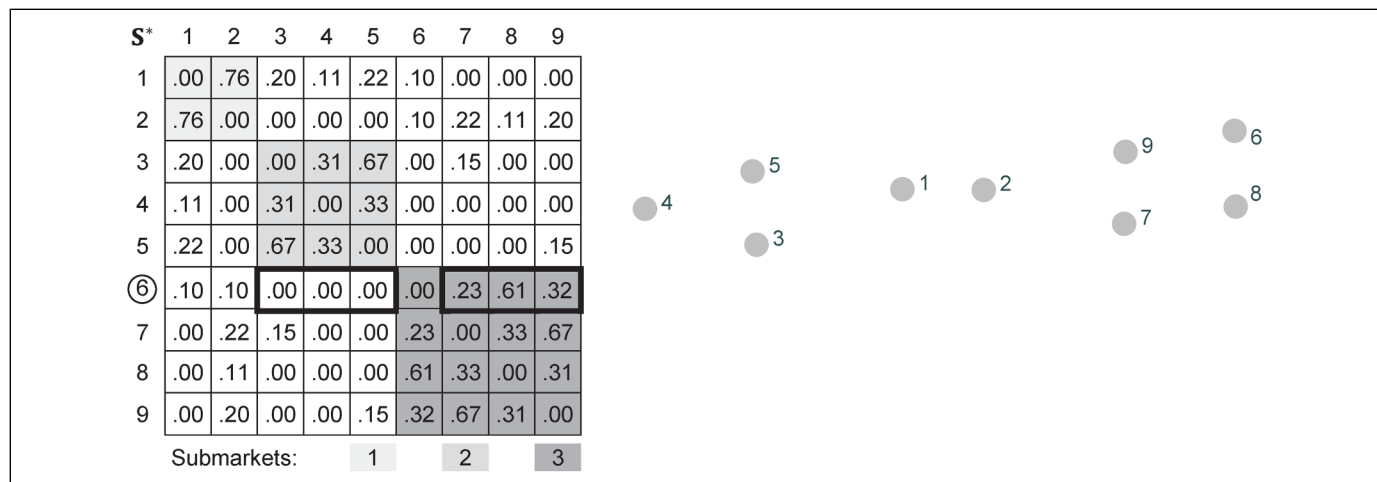


Figure 5. M4 Map for Numerical Example when Product 6 Competes in a Single Submarket.

Notes: Bubbles represent products. Labels identify products in relationship matrix S^* . Axes are on the same scale.

In addition, product attributes were collected from the website's database to explore M4's map and test it for face validity.

Inferring Competitive Relationships

Ringel and Skiera (2016) construct consideration sets from clickstream data and infer competitive relationships from products' co-occurrences in these sets. Although this approach is straightforward, it ignores that a clickstream represents a consumer's search over time. All temporal information is lost when a clickstream is collapsed to a set. Intuitively, however, products searched together in temporal proximity may be more strongly related than products searched further apart in time. Such possibility arises, for example, when consumers learn about new products or new attributes during their searches such that they update their preferences and change their product search space (Branco, Sun, and Villas-Boas 2012; Bronnenberg, Kim, and Mela 2016; Fershtman and Pavan 2020; Gardete and Hunter 2020; Greminger 2022; Ursu, Wang, and Chintagunta 2020). Thus, a different approach, namely, neural product embedding, is employed in this study.

Neural product embedding. To consider temporal information in the analysis, competitive relationships among products were inferred from clickstream data using word2vec with skip-gram⁵ by Mikolov et al. (2013). Originally developed for natural language processing (NLP) tasks, word2vec embeds words from a text corpus into high-dimensional vector space such that vectors of words that commonly share contexts in the corpus are located in proximity to one another. In contrast to autoencoders, which train against the input through reconstruction, word2vec trains input words against neighboring words in the sentences of a

text corpus. The skip-gram architecture thereby receives a word as input and predicts its context, that is, neighboring words, as output. The context is a window around a focal word containing the words that appear before and after it. Word sequence is observed as the context window "slides" across sentences during training. The embedding is updated on the basis of the prediction error when a focal word's vector cannot accurately predict its context. As the training process proceeds, the numerical vectors of similar words are nudged closer together in high-dimensional vector space.

Applied to clickstream data, word2vec creates a product embedding that is based on consumers' latent preferences for products' attributes, as expressed by their searches. A clickstream can thereby be thought of as a story about a consumer's search, written by the consumer using products instead of words. Because word2vec does not consider entire clickstreams but rather search sequences within clickstreams (i.e., context windows around focal products), this approach accounts for the possibility that products searched in temporal proximity might be more strongly related than those searched further apart in time. Thus, clickstreams were treated as a temporal continuum without imposing any additional structure on the search data, such as breaking clickstreams up into sessions by rules that require additional assumptions on the part of the analyst.

The ability of word2vec to identify substitutive relationships among products was recently shown by Gabel, Guhl, and Klapper (2019), who applied word2vec to consumers' shopping baskets to infer and visualize the assortment structure of a grocer. The advantage of such an approach is that it is based on latent product attributes derived from consumers' behaviors, which eliminates the need to make a priori assumptions on which product attributes to select for the analysis, which in turn minimizes the risk of omitting attributes that might be important to consumers or developing biased views of the market structure (Gabel, Guhl, and Klapper 2019). In contrast to the analysis of shopping baskets, where the sequence of

⁵ Embedded to 30 dimensions using a softmax classifier, a context window of 2, and a subsampling threshold of 10^{-2} .

product selection is unobserved, this study exploits word2vec's ability to explicitly consider temporal information (i.e., search sequence via its context window) in consumers' clickstreams to infer competitive relationships among products.

Noise in search data and stochasticity in neural product embedding. Two more challenges were addressed before the analysis was conducted. First, search data can be noisy. Second, neural product embedding is stochastic in nature, such that multiple trainings on the same data may not produce identical results (Antoniak and Mimno 2018). Both can compromise the robustness of the empirical findings. Thus, a bootstrapping procedure was employed to jointly address both challenges and arrive at a robust solution. Specifically, the procedure is to (1) bootstrap the collected clickstreams with replacement, (2) train a neural product embedding on each bootstrap, (3) generate a weighted graph (i.e., weighted pairwise product relationships) from each bootstrap's embedding, and (4) average products' weighted relationships across bootstraps to obtain the relationship matrix (\mathbf{S}).

Steps 3 and 4 of this bootstrapping procedure were necessary because a single relationship matrix \mathbf{S} cannot be derived directly from multiple (bootstrapped) product embeddings. The reason is that product embeddings are not directly comparable. Although the same notions of substructures may live in each of multiple product embeddings, these have different meanings with respect to each embedding's underlying geometry. Put more simply, the individual (latent) dimensions of multiple product embeddings are not necessarily aligned such that they can be directly compared across embeddings.

To resolve this problem, a weighted graph (i.e., products' weighted pairwise relationships in high-dimensional space) was first generated from each product embedding using the first phase of a nonlinear manifold approximation and projection technique called UMAP⁶ (McInnes and Healy 2018). Because products' pairwise relationships in high-dimensional space are now unambiguously identified for each bootstrap, they can be averaged across bootstraps to obtain a robust system of product relationships (i.e., relationship matrix \mathbf{S}).

UMAP's first phase (uniform manifold approximation) is rooted in the idea of constructing simplicial complexes, that is, Vietoris–Rips complexes (Hausmann 1995), from topological spaces to generate a weighted graph in high-dimensional space. This weighted graph is essentially a single fuzzy simplicial complex constructed from the union of fuzzy simplicial sets. The relevant members of each fuzzy set (i.e., its objects) are the neighbors of any given object, whereby fuzzy set

membership strengths decay rapidly beyond the set perimeter. UMAP was used in the bootstrapping procedure because it automatically adjusts distance scales according to changing densities of a manifold, which enables it to more accurately capture both global (market) and local (submarket) structures. UMAP's second phase (projection) was not used in this study because it was found to produce inferior results relative to t-SNE's map projection for the DSLR camera market (see Web Appendix E for details).

Accommodating Consumer Heterogeneity

SpeakEasy and t-SNE, which are used inside the M4 framework, are not explicitly designed to model consumer heterogeneity. Nonetheless, the approach for inferring competitive relationships in the DSLR camera market implicitly accommodates consumer heterogeneity. In the clickstream data, each consumer searches only a small subspace of the entire product space on the basis of their preferences. Because of heterogeneity in consumers' preferences, these subspaces are not identical across consumers. Accordingly, each consumer in the data only contributes information to relationship matrix \mathbf{S} for a fraction of all products, namely those products that they actually searched for. Submarkets emerge when many consumers share search spaces that are different from the shared search spaces of other consumers. Put differently, heterogeneity in consumers' preferences, which are embedded in their searches, drives varying relationship densities in the neural product embeddings. These varying relationship densities in higher-dimensional space emerge as submarket structures on the map of the DSLR camera market. Indeed, this study shows that consumers' clickstreams are strongly concentrated in the submarkets on M4's market structure map (see the "Map Validity" subsection for details).

The advantage of this approach is that it generates a single, comprehensive market structure map from many different consumers. Individual competitive relationships are identified from only those consumers who actually shop for the products that pertain to these relationships. In this regard, this approach is fundamentally different from approaches that elicit all pairwise product relationships in a market from all respondents of a study. Particularly, there is no need to make two strong assumptions, namely (1) that all consumers are aware of all products in a market and (2) that all consumers know the vertical and horizontal distributions of products' attributes.

Nevertheless, the implicit accommodation of consumer heterogeneity is not without limitation. First, heterogeneous consumer behavior might not be directly accessible in alternative types of data that relationship matrix \mathbf{S} could be constructed from. Second, the proposed approach does not allow statistical inference of the preferences that drive consumers' searches in each submarket. And, third, it is possible that some within-submarket heterogeneity remains on the maps (cf. Franke, Reisinger, and Hoppe 2009). Although these limitations do not invalidate the proposed approach, analysts need to be

⁶ UMAP (uniform manifold approximation and projection) comprises two phases. The first phase (uniform manifold approximation) generates a high-dimensional weighted graph (i.e., a fuzzy topological representation of product's relationships in high-dimensional space). The second phase (projection) projects the weighted graph onto a lower-dimensional space. Only the first phase is used in the bootstrapping procedure.

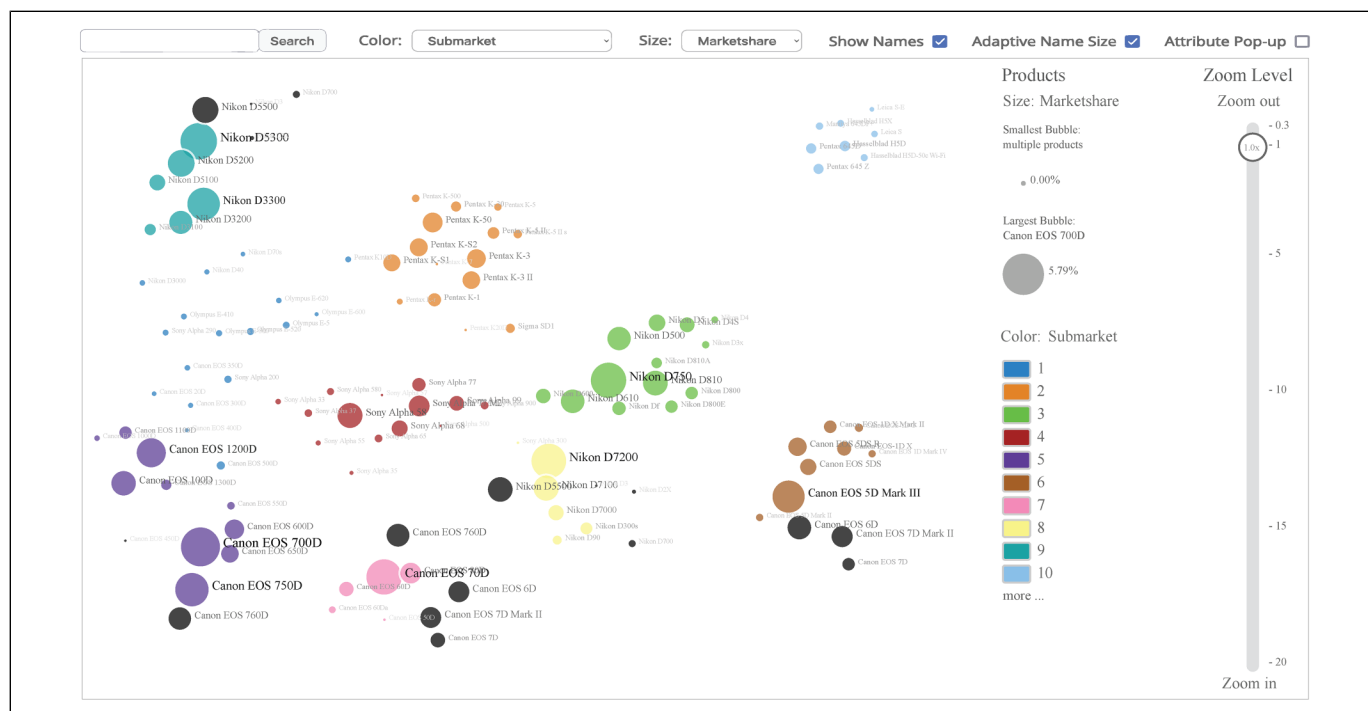


Figure 6. Multimarket Membership Map of the DSLR Camera Market Visualized in mapXP.

Notes: Axes are on the same scale. MMPs (indicated in black) are Nikon D5500, Canon EOS 760D, Nikon D3, Canon EOS 7D, Canon EOS 450D, Nikon D2X, Canon EOS 6D, Canon EOS 7D Mark II, and Nikon D700.

aware of them and consider them in their interpretation of the entailing market structure maps.

Findings

M4 identified ten submarkets and nine MMPs, that is, cameras that compete in the context of more than one of the identified submarkets. Half of all submarkets are affected by MMPs. Figure 6 shows the market structure map of the DSLR camera market obtained with M4 (see Web Appendix F for enlarged maps). Although the map's two dimensions are latent and thus do not have a global meaning, it is possible to discern drivers of product similarity within and across individual submarkets using ensemble coding that overlays product attributes onto the map by varying bubble color, and labels (Ringel and Skiera 2016). To this end, an interactive online tool called mapXP allows readers to easily explore all maps of this study (www.mapxp.app/DSLR).

Here, mapXP is used to explore M4's map. As indicated by bubble color in Figure 6, the submarkets that SpeakEasy identifies are well recovered in M4's map. The most popular DSLR cameras in terms of product views (bubble size in Figure 6) are distributed across the various submarkets where they compete in specific local contexts against other cameras. A closer look at Figure 6 reveals that all identified MMPs (black bubbles) are either Canon or Nikon cameras, which warrants a deeper investigation into how these products are positioned by their manufacturers. Manufacturers of DSLR cameras generally

position each camera within one of three product tiers, as indicated in the camera's product description, namely "standard," "semi-professional," and "professional" cameras. This split into three product tiers is consistent with the consumer search patterns on which M4's map is based: both Canon and Nikon cameras are largely distributed across three submarkets in M4's map.

By overlaying manufacturers' positionings over M4's map in Figure 7, one can explore the extent to which manufacturers' positionings are consistent with consumers' perceptions of viable substitute cameras. Overall, cameras of the same product tier are indeed found close to one another on M4's map (as indicated by bubble color in Figure 7). There are, however, exceptions. Most notably, the identified MMPs (labeled with their names in Figure 7) tend to compete across multiple product tiers. For instance, the Canon EOS 760D and EOS 7D Mark II (circled in Figure 7) are positioned by the manufacturer in a lower product tier but are evidently also considered by consumers searching for cameras of a higher product tier.

Assuming higher absolute profits for higher-tier products, this mismatch warrants a deeper investigation into the positioning and pricing of the EOS 760D and EOS 7D Mark II since they may be cannibalizing the sales of Canon's higher-tier cameras. At the same time, these cameras may nudge consumers searching for a lower-tier camera toward considering Canon's higher-tier cameras, which would be favorable to Canon.

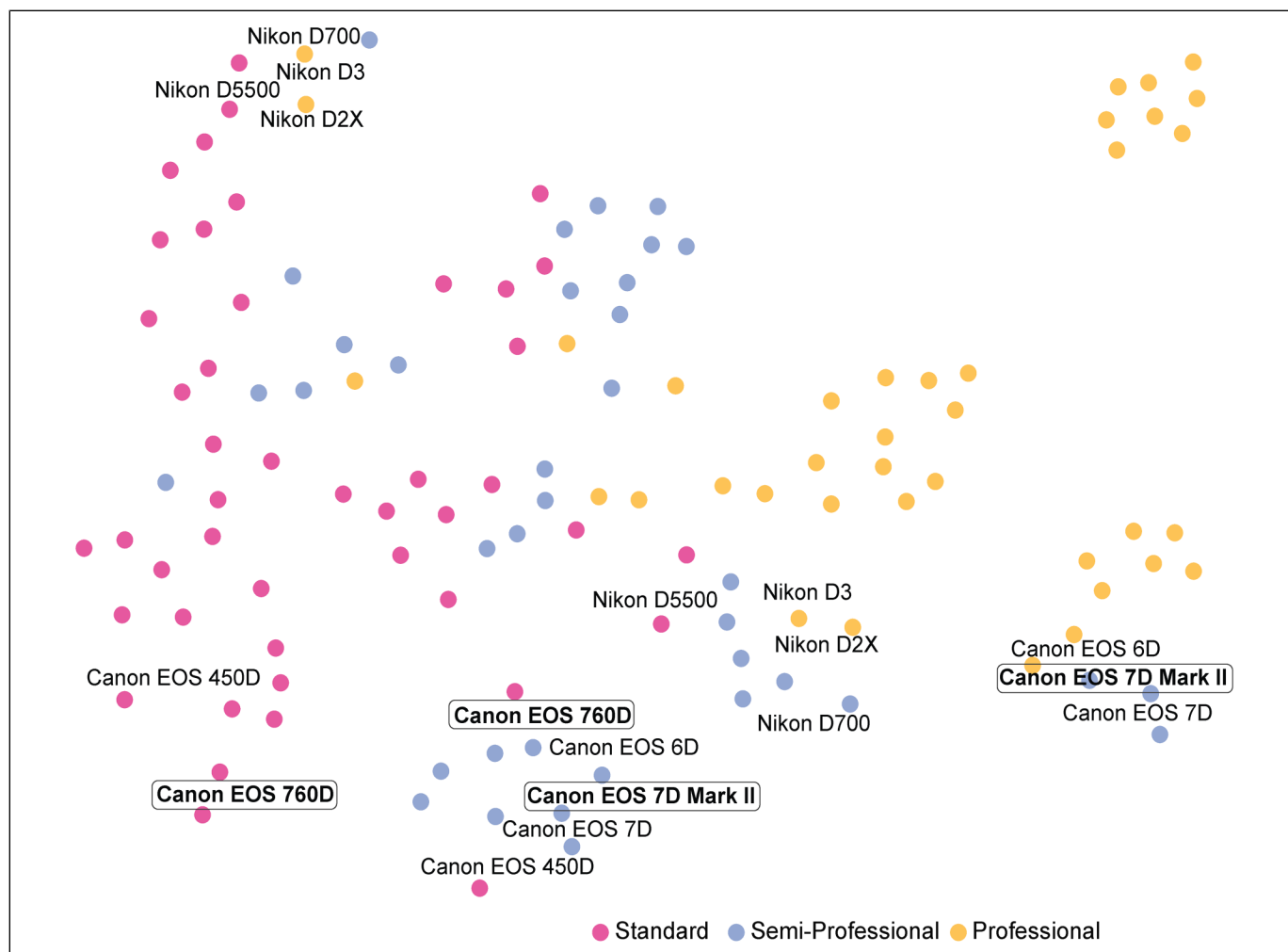


Figure 7. Manufacturers' Positioning by Product Tier in the DSLR Camera Market.

Notes: Bubbles represent cameras. Axes are on the same scale. Bubble color indicates manufacturer's positioning (standard, semi-professional, professional).

To better understand whether such nudging takes place or not, the next step was to examine the raw clickstream data, which capture consumers' actual search patterns, more closely. Specifically, the aim was to investigate whether consumers who view a multimarket membership camera confine their search to only one of the submarkets the camera competes in, or whether consumers continue their search in the respective other submarket after viewing the camera (i.e., transition from one submarket to a different submarket). For Canon's EOS 760D and 7D Mark II cameras, the findings show that consumers largely confine their searches to either one or the other of these cameras' submarkets (see Table 2). Switching between submarkets takes place less frequently with little difference in terms of direction. This pattern is consistent across all identified multimarket membership cameras, with less than 25% of consumers shifting their search from one submarket to another after viewing a multimarket membership camera.

Overall, model-free evidence from consumers' underlying search patterns suggests that the MMPs identified by M4 do

not serve as transition products that nudge consumers to higher-tier cameras. Instead, the identified MMPs point to potentially harmful inconsistencies in manufacturers' positioning strategies that may need to be revisited. A deeper investigation should also include a financial evaluation of MMPs' net impact on firm profit to determine whether MMPs affect total profit positively despite some (possibly unintended) substitution.

Map Validity

Arguably, market structure maps are only valuable to managers when the competitive relationships visualized therein are valid. Thus, the approach for inferring competitive relationships from search data, and the entailing market structure map, were tested for their validity.

Although the idea to derive competitive relationships from search data on the basis of co-consideration of products has been laid out conceptionally (e.g., Kim, Albuquerque, and

Table 2. Search Patterns Across Submarkets for Multimarket Membership Cameras.

Search Patterns Across Submarkets	Percentage Share of Searches	
	Canon EOS 760D	Canon EOS 7D Mark II
Search in Submarket 5, not 7	48.12	—
Search in Submarket 7, not 5	24.72	—
Start searching in Submarket 7, view multimarket membership camera, continue in Submarket 5	9.80	—
Start searching in Submarket 5, view multimarket membership camera, continue in Submarket 7	8.56	—
Switch back and forth between Submarkets 5 and 7	8.80	—
Search in Submarket 6, not Submarket 7	—	39.92
Search in Submarket 7, not Submarket 6	—	38.56
Start searching in Submarket 6, view multimarket membership camera, continue in Submarket 7	—	7.81
Start searching in Submarket 7, view multimarket membership camera, continue in Submarket 6	—	6.00
Switch back and forth between Submarkets 7 and 6	—	7.71

Notes: To observe submarket switching, the analysis only considers consumers who searched at least three cameras. Submarket numbers are as shown in Figure 6: Submarket 5 (purple); Submarket 6 (brown); Submarket 7 (pink).

Bronnenberg 2011; Moe 2006; Ringel and Skiera 2016), no additional evidence in support of this idea is available. To close this gap, the premise that competitive relationships can be inferred from consumers' searches was tested in a survey of 500 consumers about their use of price comparison sites. The objective was to determine (1) whether consumers actually consider the products for purchase that they search at price comparison sites and (2) when during their decision process consumers use price comparison sites. The survey comprised 15 questions, which are available in Web Appendix G. Respondents were recruited anonymously from a commercial research provider's participant pool. To participate in the study, respondents had to be age 18 years or older and must have shopped online before.

The findings show that over 92% of the price comparison site users in the survey use these sites to view products that they are considering or intend to purchase. Less than 1% view products that they do not consider purchasing, and less than 7% view products that they are only generally interested in. These findings support the idea that consumers' searches at price comparison sites reveal co-consideration and can thus provide insights into competitive relationships. Further, most price comparison site users (56%) visit price comparison sites at different points during their decision process, which supports the approach of considering the temporal proximity of jointly

searched products in consumers' clickstreams using neural product embedding.

The next step was to examine how well M4's map recovers consumers' actual search patterns in the underlying raw clickstream data. Specifically, two concentration metrics—the Hirschman–Herfindahl Index (HHI)⁷ and entropy⁸—were calculated to test the extent to which consumers' actual searches are concentrated within the submarkets that are visualized in M4's map. The results show a mean HHI of .81 and a mean entropy of .3. Both metrics indicate that consumers' searches are strongly, but not entirely, concentrated in M4's submarkets, which supports the internal validity of M4's map.

The final step was to take a closer look at M4's map to determine whether the visualized market structure can be intuitively explained by camera attributes overlaid onto the map using mapXP. The findings show that M4's map exhibits clear submarket structures that appear to be driven by brand (see the bubble labels in Figure 6). This hierarchy of brand over product is in line with the findings of Voleti, Kopalle, and Ghosh (2015), who modeled brand and product competition using panel data. The pattern may also reflect the high switching costs that consumers who already own a DSLR camera would incur if they moved to a new brand, owing to incompatibility of accessories across brands. The findings further show that manufacturers' positionings (across the three product tiers that were extracted from their product descriptions) are largely recovered in M4's map (Figure 7), with cameras positioned as standard DSLR cameras (fuchsia bubbles) located toward the left, whereas cameras positioned as professional DSLR cameras (gold bubbles) are located toward the right.

The identified market structure also appears to be driven by megapixel and price, as shown in Figure 8, where both consistently increase from left to right and are largely correlated ($r = .596$). This correlation is to be expected, given that higher picture quality should come at a higher price. There are, however, exceptions, notably among the MMPs. A closer look at Canon's EOS 760D and 7D Mark II cameras reveals that both have attribute levels similar to those of cameras of the subsequent (higher) product tier—for which they were not positioned by their manufacturers (see Figure 8). For instance, the EOS 7D Mark II features a high-resolution sensor (i.e., high megapixels) and a solid camera body (indicated by camera weight) that are both typical for professional cameras. Yet, Canon's EOS 7D Mark II is positioned as a semi-professional camera and priced accordingly at a lower price point than Canon's professional cameras. Thus, the EOS 7D Mark II creates different benefits for consumers shopping for a semi-professional camera (where it is a premium option) than for consumers shopping for a professional camera (where it represents a bargain). The camera's multimarket membership is thus

⁷ $HHI = \sum_{i=1}^n (MS_i)^2$, where $i \in n$ submarkets with which the products of a consumer's clickstream are associated, and MS_i corresponds to the share of submarket i of a consumer's clickstream.

⁸ Entropy $= - \sum_{i=1}^n P(x_i) \log P(x_i)$, where $i \in n$ submarkets with which the products of a consumer's clickstream are associated, and $P(x_i)$ corresponds to the share of each submarket of a consumer's clickstream.

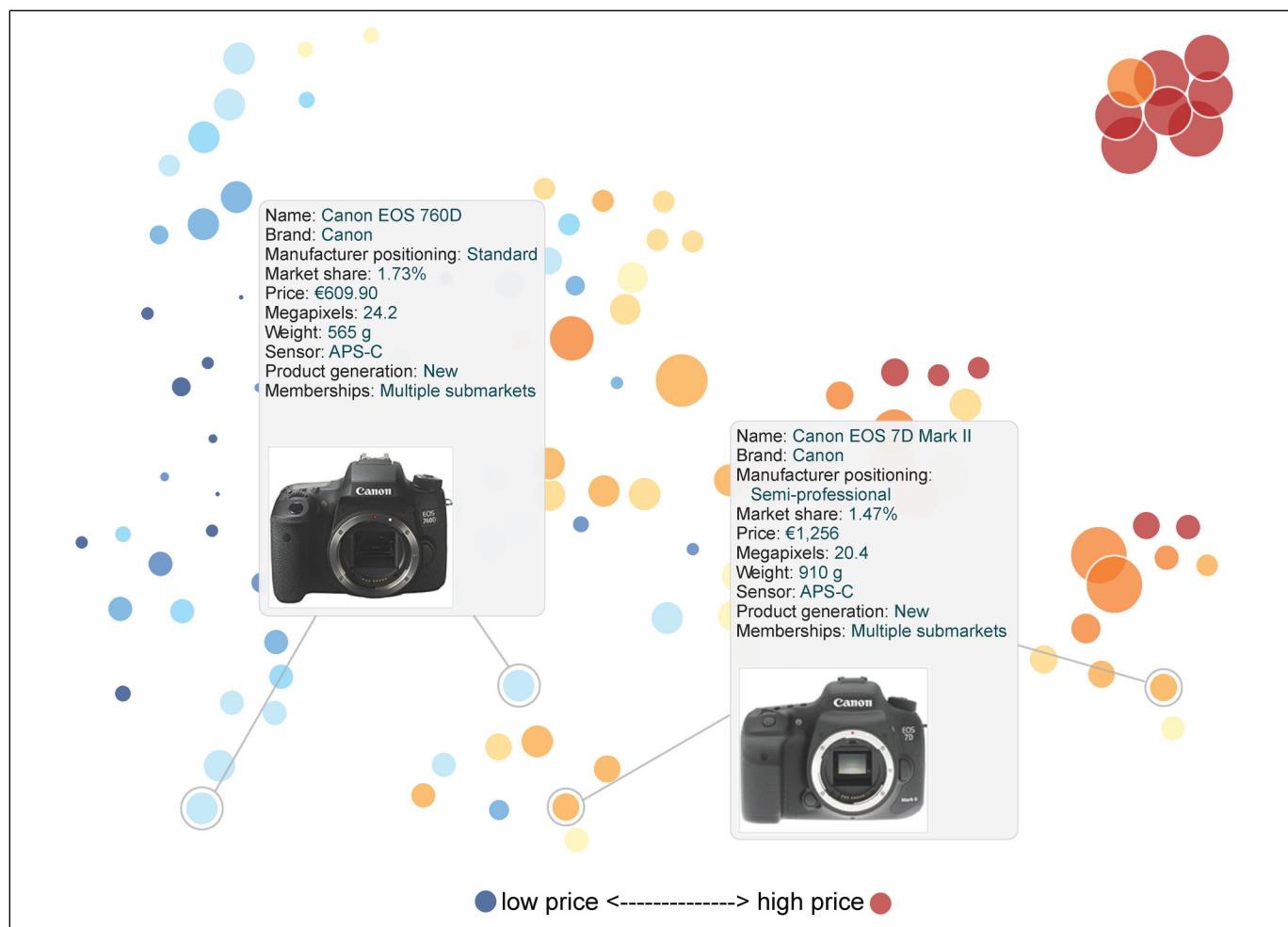


Figure 8. Price and Megapixels in the DSLR Camera Market.

Notes: Bubbles represent cameras. Axes are on the same scale. Bubble size indicates megapixels. Bubble color indicates price (heat map).

plausible from consumers' perspectives. It is further supported by consumers' actual search patterns, which are largely confined to either of the two submarkets (compare Table 2).

Taken together, these findings support the approach for inferring competitive relationships from online search data and the validity of M4's map: the visualized market structure appears face valid, it can be related back to the underlying clickstreams by model-free evidence, and the identified multimarket memberships can be rationally explained by products' attributes.

Robustness Tests

Several robustness tests were conducted to better understand the extent to which the market structure inferred using M4 is sensitive to variations in the underlying search data. Specifically, the testing involved a tenfold cross-validation in which 10% of the observed consumers were randomly removed from the raw clickstream data of each fold. Three different metrics were then used to systematically test for robustness at the global, local, and individual product levels.

First, the Pearson correlation among all pairs of inferred relationship matrices was calculated to determine how similar the global market structure is across folds, resulting in a mean correlation of .987. Next, to investigate how similar local submarket memberships of cameras are across folds, the Omega index⁹ (Collins and Dent 1988) was used. The Omega index builds on both the Rand index and the adjusted Rand index by accounting for disjoint solutions and correcting for chance agreement. It

⁹ Omega (Ω) for two clustering solutions s_1, s_2 is calculated from observed agreement (O) and expected agreement (E) by $\Omega(s_1, s_2) = (O(s_1, s_2) - E(s_1, s_2)) / (1 - E(s_1, s_2))$. Observed agreement is an extension of the Rand index given by $O(s_1, s_2) = \sum_{j=0}^{\min(J,K)} A_j / N$, where J and K represent the maximum number of submarkets in which any pair of products appears together in solutions 1 and 2, respectively; A_j is the number of the pairs agreed by both solutions to be assigned to number of submarkets j ; and N is the number of product pairs. Expected agreement is based on the adjusted Rand index and is given by $E(s_1, s_2) = \sum_{j=0}^{\min(J,K)} N_{j1} N_{j2} / N^2$, where N_{j1} is the total number of product pairs assigned to number of submarkets j in solution 1, and N_{j2} is the total number of product pairs assigned to number of submarkets j in solution 2.

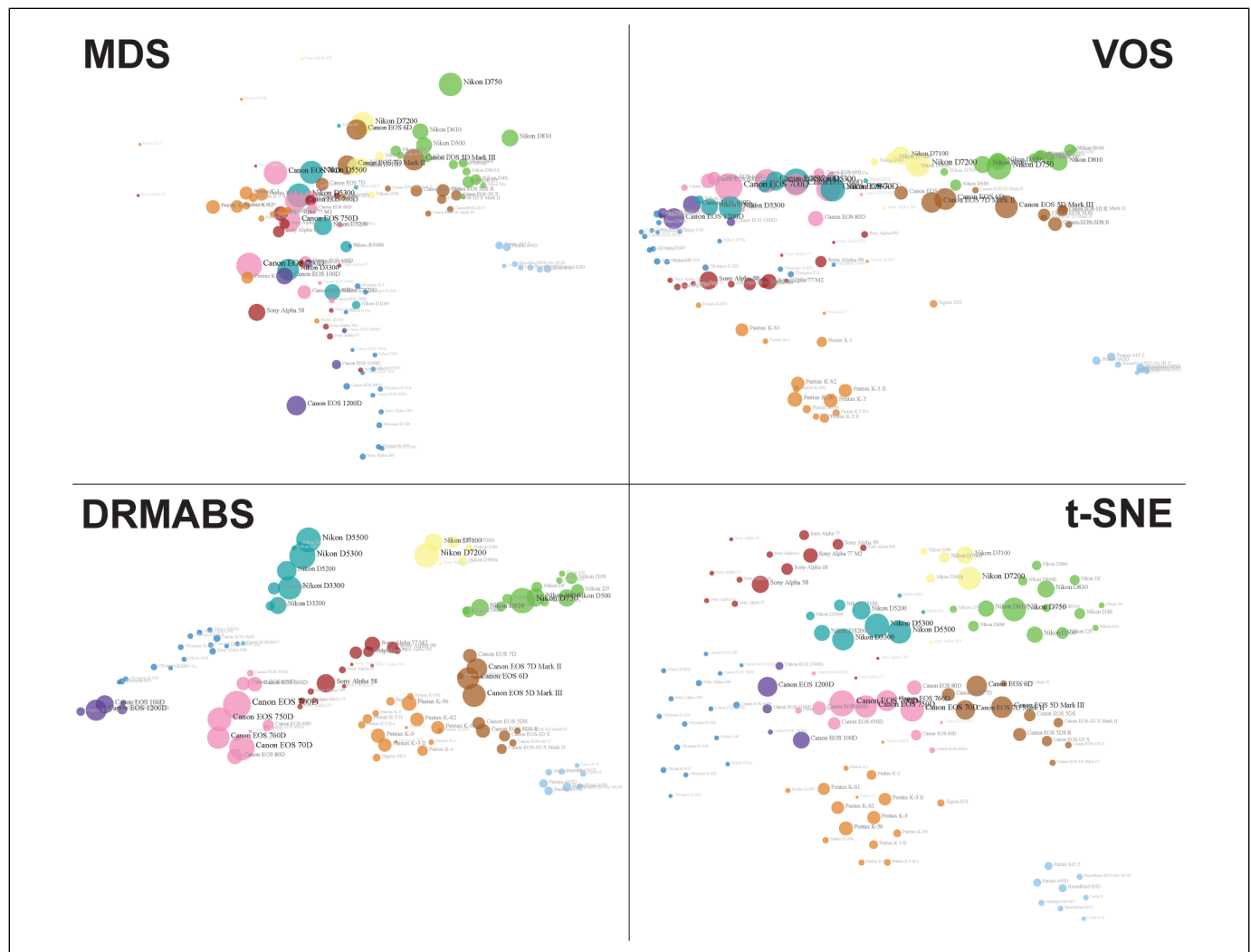


Figure 9. Alternative Methods' Maps for the DSLR Camera Market.

Notes: Bubbles represent cameras. Bubble color indicates submarket. Bubble size is based on number of camera views. Axes are on the same scale. For illustrative purposes, Louvain community detection (Blondel et al. 2008) was used to identify submarkets for the MDS, VOS, and t-SNE maps because these methods do not explicitly do so themselves.

ranges from 0 (no agreement) to 1 (perfect agreement). The mean Omega value of .93 indicates strong agreement. Finally, the extent to which cameras face the same closest competitors across folds was tested. To do so, a metric called k-CC agreement was introduced. This metric measures the mean share of the number of corresponding k closest competitors for each camera between pairs of folds. The mean k-CC agreement of 91% (for k=5) shows that cameras largely face the same closest competitors across all ten folds. Overall, these results suggest that the inferred market structure is relatively robust to some variation in the underlying search data.

Method Comparison

To better understand the benefits of using M4 in an empirical setting, its market structure map of the DSLR camera market was compared with those of alternative mapping methods

(shown in Figure 9). Each market structure map was evaluated in four ways. First, a hit rate (sometimes also referred to as recovery rate) was used to determine how accurately each map represents local competitive relationships among individual products. Second, the accuracy of each map was further qualified with the use of error-weighted classification metrics (i.e., precision, recall, and F_1 score). Third, each map was examined to determine how well it visually recovers the global structure of the market (i.e., its submarkets). Fourth, each map was visually inspected for face validity.

To determine how accurate local competitive relationships are represented in maps, this research uses a metric that is independent of mapping methods' objective functions but relates directly to the objective of visualizing market structure, namely, to position products on a map in such a way that they are located close to their closest competitors (Faure and Natter 2010). Specifically, this approach follows recent research

Table 3. Method Comparison in the DSLR Camera Market.

Method	Hit Rates as Percentage			Error-Weighted Classification Metrics			Visual Recovery of Global Structure (Omega)
	All	MMPs	Non-MMPs	Precision	Recall	F ₁ score	
MDS	26.12	13.33	27.25	.23	.49	.31	.32
VOS	42.88	28.88	44.11	.40	.62	.49	.46
DRMABS	62.52	51.11	63.53	.60	.79	.68	.69
t-SNE	62.16	55.55	62.74	.62	.79	.69	.67
M4	68.31	64.44	68.65	.67	.82	.74	.84

Notes: Error-weighted classification metrics range from 0 (worst) to 1 (best). Omega ranges from 0 (no visual recovery of global structure) to 1 (perfect visual recovery of global structure). Hit rates are robust across the top one, three, and five closest competitors.

in market structure visualization that uses a hit rate¹⁰ to measure map accuracy (e.g., Gabel, Guhl, and Klapper 2019; Ringel and Skiera 2016). In this context, a product's hit rate is the share of its top five closest competitors that are among the five products most proximal to it on the map, which is a special case of the rank order agreement by Akkucuk and Carroll (2006). The notion of the utilized hit rate is in line with Waldo Tobler's first law of geography: "Everything is related to everything else, but nearby things are more related than distant things" (Tobler 1970, p. 236). Likewise, according to construal level theory (Trope and Liberman 2010), the more distant an object is from the individual, the more abstractly it will be thought of, whereas closer objects are thought of more concretely. Three different hit rates are used to compare the maps of this study: for all products, for MMPs, and for non-MMPs. The findings show that the M4 framework outperforms all alternative methods on all three hit rates (see Table 3).

To obtain a more nuanced perspective on map accuracy, an approach was borrowed from the performance evaluation of classification models in supervised learning and modified to operate in the context of market structure maps. Specifically, an error-weighted confusion matrix was generated and used to calculate the precision, recall, and F₁ score for each map in the comparison. Whereas precision considers the errors related to incorrectly identified competitors, recall considers the errors pertaining to unidentified competitors. The F₁ score is the harmonic mean of the two metrics.¹¹

The prerequisite for a confusion matrix is a classification task for which a method's prediction can be compared with the truth. For the task of accurately projecting products' competitive relationships onto a map, prediction was defined as the products closest to a focal product on the map, and truth was defined as the competitors closest to the focal product in the underlying data. Then, the products that are closest to a

given focal product on a map (in terms of Euclidean distance) are either its closest competitors in the underlying data (true positive [TP]) or are not its closest competitors in the underlying data (false positive [FP]). Further, the products that are not closest to a given focal product on a map are either not its closest competitors in the underlying data (true negative [TN]) or are its closest competitors in the underlying data (false negative [FN]).

In addition, the confusion matrix was weighted by the magnitude of the error between prediction and truth. Specifically, TP, FP, FN, and TN were weighted by the absolute deviations between products' proximities on the map and their relationship strengths in the underlying relationship matrix such that smaller deviations contribute more to TP and TN (i.e., reward), and larger deviations contribute more to FP and FN (i.e., penalty). The introduced weighting scheme is in line with the objective of classification tasks: TP and TN should be as high as possible, whereas FP and FN should be as low as possible. To obtain absolute deviations between map proximities (*p*) and relationship strengths (*s*), the data are normalized to the same range [0, 1], and Euclidean map distances (*d*) are converted to proximities (i.e., $p = 1 - d$). Then, for every product in the market, (1) the sum of the absolute deviations $|p - s|$ to the product's closest competitors is calculated for FP and FN, (2) the sum of the inverted absolute deviations $1 - |p - s|$ to the product's closest competitors is calculated for TP and TN, and (3) the two sums are added to the respective totals in the error-weighted confusion matrix. As shown in Table 3, the M4 framework outperforms alternative methods on all three performance metrics (i.e., precision, recall, F₁ score) based on the products' five closest competitors for each map.

The next analysis in the method comparison was an investigation of how well the generated maps visually recover global market structures (i.e., submarkets). The premise is that the identified submarkets should show up as "groups" or "clusters" of products on the generated maps so that managers can discern them and draw correct conclusions. To determine to what extent this is the case, the Omega index is used to test how strongly the visualized submarket structures resemble the submarkets that were detected in the underlying relationship data (i.e., relationship matrix *S*). The products' map coordinates are clustered with *k*-means clustering to identify submarket structures on

¹⁰ To calculate hit rate (HR), let N_X and N_Y be two neighborhood matrices where, for every product $i \in n$, its j nearest neighbors are captured in higher-dimensional space (i.e., relationship matrix *S*) and on a map projection (in terms of Euclidean distance), respectively. Then, $HR_j = \frac{1}{j} \sum_{i=1}^j a_{ij}$, where a_{ij} is the number of products shared by N_X and N_Y .

¹¹ $F_1 \text{ score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$, where $\text{Precision} = \frac{TP}{TP + FP}$ and $\text{Recall} = \frac{TP}{TP + FN}$.

maps. The result shows that M4's map visually recovers global market structures best, followed by DRMABS and t-SNE (see Table 3).

Finally, all maps were visually inspected for their face validity. The MDS map in Figure 9 exhibits circular bending and lumping of products, a known weakness of classic MDS, particularly in larger markets. Accordingly, the submarket structures are poorly recovered. Although submarkets are more discernable in the VOS map, they appear to be stretched, with several cameras loosely distributed across the map center. The DRMABS and t-SNE maps exhibit much clearer submarket structures than those of MDS and VOS. However, from a face validity perspective, it is difficult to explain why, in both maps, Submarket 10 (light blue), which consists of top-of-the-line professional cameras, is separated from Nikon's flagship cameras in Submarket 3 (green), although it does appear in the vicinity of Canon's flagship cameras in Submarket 6 (brown). Given that experts generally consider Nikon's flagship cameras as being on par with Canon's flagship cameras, one would expect the top-of-the-line professional cameras of Submarket 10 (light blue) to be in the vicinity of both brands' flagship cameras. Moreover, Submarket 4 (red), which contains the bulk of Sony's cameras, is pushed off to the top in t-SNE's map near Nikon's cameras, whereas it is located near Canon's cameras in the DRMABS map. Given that Sony has consistently built its presence in the DSLR camera market, where it offers several standard and semi-professional cameras at attractive prices, it seems implausible that Sony's cameras would only immediately compete either with Nikon cameras (as in the t-SNE map) or with Canon cameras (as in the DRMABS map).

The M4 map not only performs best across all metrics in Table 3 but also resolves the structural implausibilities of the other methods' maps: Submarket 10 (light blue), with its top-of-the-line professional cameras, is close to both Canon's and Nikon's flagship cameras (Figure 6). Sony's cameras in Submarket 4 (red) are located toward the center of the M4 map, where they compete with cameras of multiple brands (Figure 6).

The visualization in Figure 10 reveals more intricate differences between the t-SNE map and the M4 map of the DSLR camera market. The heat map in this figure shows where cameras' five nearest neighbors differ between maps (i.e., local differences), and circle submarkets that appear in different global map positions. Beyond differences in the global structure (black ellipses) and for MMPs (largely marked in orange, that is, approximately 50% agreement), the figure also reveals more subtle differences for other cameras (largely marked in light blue color, that is, approximately 70% agreement).

Overall, the findings of the method comparison for the DSLR camera market are consistent with those of the numerical example: Beyond identifying MMPs, M4 visualizes both local and global competitive relationship structures more accurately than alternative methods do. Notably, the presence of MMPs also affects how relationships among other

products are visually represented on maps, and how accurate these representations are.

Simulation Study

The numerical example and the empirical study of the DSLR camera market both demonstrate the benefits of accounting for MMPs in market structure maps. However, markets may differ in regard to the number of submarkets that constitute them and the number of MMPs that compete across these submarkets. A simulation study was therefore run to compare the maps produced by the M4 framework with those of alternative methods across various market structures.

The LFR graph benchmark by Lancichinetti, Fortunato, and Radicchi (2008) was used to simulate 2,400 unique market structures (i.e., matrices S) of 100 products that exhibit real-world properties such as heterogeneity in the distribution of relationship strengths and variation in submarket sizes. The simulation comprises varying numbers of submarkets ($\mu = 6.36$, $\sigma = 1.68$) with few (5%), moderate (10%), and many (20%) MMPs. The objective is to better understand the extent to which the M4 framework produces better (or worse) market structure maps than alternative mapping methods in regard to (1) accurately representing local competitive relationships among individual products and (2) accurately visualizing the global structure of the market (i.e., its submarkets).

In total, the simulation study comprises 12,000 market structure maps. The maps of all mapping methods that were used in the numerical example and the empirical study of the DSLR camera market (i.e., MDS, VOS, DRMABS, and t-SNE) were compared with the maps of the M4 framework. Overall, the findings show that the M4 framework consistently produces more accurate maps in terms of hit rate than alternative mapping methods, outperforms them on error-weighted classification metrics, and visually recovers the global structure (i.e., discernable submarkets) best. Consistent with the findings of the empirical study of the DSLR camera market, MDS performs worst on local and global accuracy, followed by VOS, DRMABS, and t-SNE, the next best method after M4. Table 4 presents the percentage gain of M4 over the best alternative mapping method (t-SNE) for all metrics. The findings are statistically significant at the 99.9% level ($p < .001$ paired sample t-test).

Notably, both MMPs and non-MMPs profit from M4 in terms of their (local) hit rates. However, the gains in hit rate diminish as the number of MMPs in the market decreases. The same is true for the error-weighted classification metrics and the visual recovery of the global market structure (as captured by the Omega index), which are less impacted in the presence of few MMPs. These findings support the conclusion that (1) the visual representation of MMPs' competitive positions on market structure maps profits from the M4 framework, and (2) potentially misleading map distortions are less severe when a market has fewer MMPs.

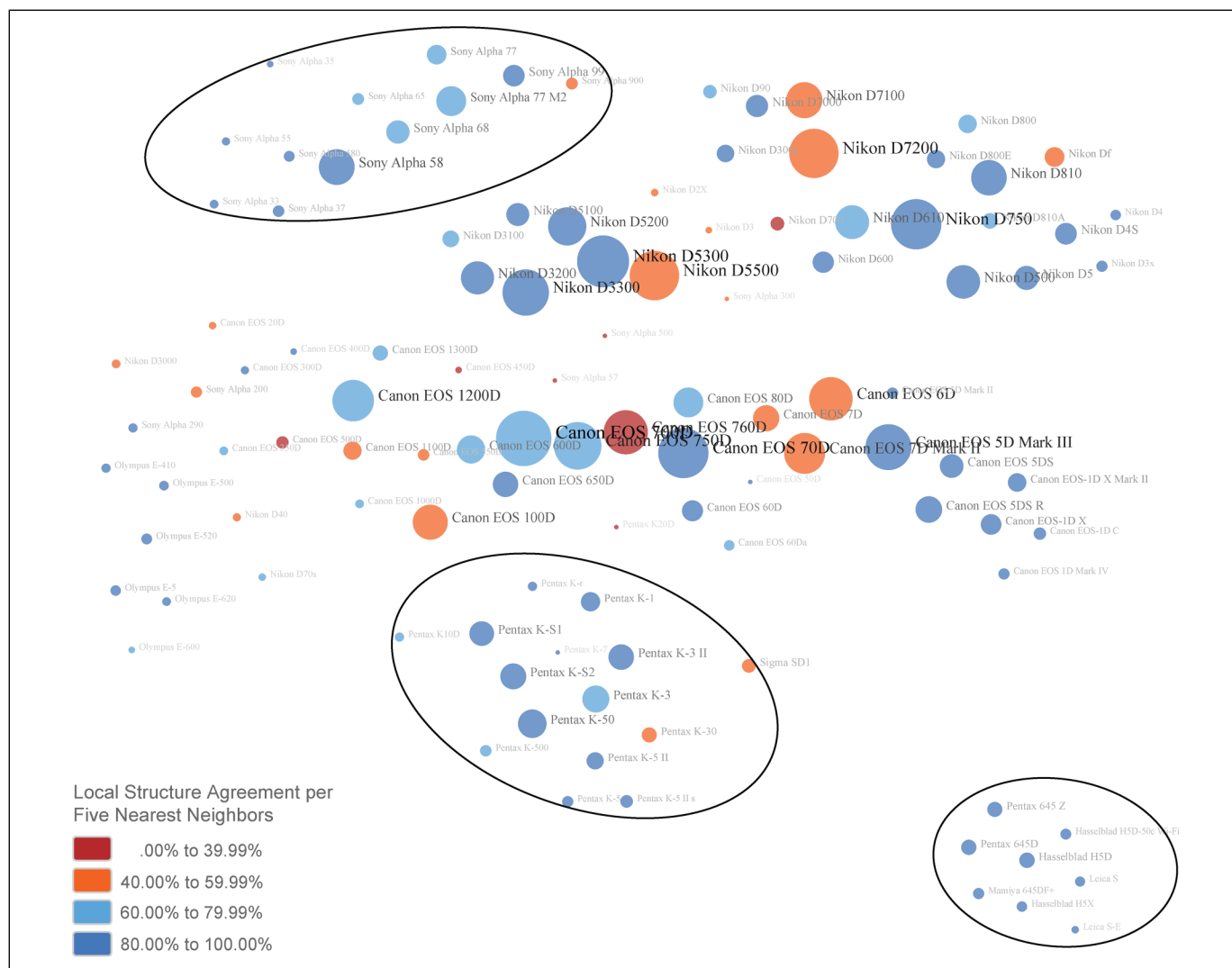


Figure 10. Differences Between t-SNE's and M4's Map of the DSLR Camera Market.
 Notes: Global structure is marked with ellipses.

Contributions to Marketing Research

This research offers both methodological and substantive contributions. On the methodological front, it contributes a novel visualization framework called M4 to market structure analysis that is not constrained to visualizing products in single map locations. At its core, the M4 framework connects overlapping clustering with map projection in a novel way—without locking the analyst into a particular clustering or mapping method. This built-in flexibility enables analysts (1) to easily tune the M4 framework to the specificities of other empirical settings and (2) to capitalize on future advances in overlapping clustering and mapping methods.

On the substantive front, this research shows that extant mapping methods are, by their design, not suitable to visualize market structure in markets where multimarket memberships of products cannot be ruled out a priori. In the presence of MMPs, their mapping outputs are subject to distortions that may mislead firms to draw wrong conclusions about the competitors

they face, the overall market structure, and the effectiveness of their positioning strategies.

Finally, while the visualization of market structure in larger markets is challenging (Gabel, Guhl, and Klapper 2019; Netzer et al. 2012; Rao and Katz 1971; Ringel and Skiera 2016), so is viewing and exploring the entailing maps. As Heer and Shneiderman (2012) point out, large and complex data sets require tools to sort, select, navigate, annotate, and share data. On that front, this article contributes an interactive map exploration tool called mapXP for use in research and practice. All maps presented in this article can be viewed and interactively explored in a web browser using mapXP (visit www.mapxp.app).

Discussion and Future Research Directions

Visualization of market structures is a valuable means for data discovery and is integral to modern business analytics (Gartner

Table 4. Percentage Increase of M4 over t-SNE in Simulation Study.

Simulation	N	Percentage Increase in Hit Rates			Percentage Increase in Error-Weighted Classification Metrics			Percentage Increase in Visual Recovery of Global Structure (Omega)
		All	MMPs	Non-MMPs	Precision	Recall	F1 Score	
Overall	2,400	10.52 (.252)	95.86 (2.408)	6.82 (.212)	5.22 (.177)	.69 (.120)	3.24 (.130)	7.56 (.166)
Few MMPs (5%)	800	2.86 (.212)	105.75 (6.038)	1.39 (.163)	1.20 (.162)	-.02 (.128) ^{NS}	.65 (.126)	2.07 (.090)
Moderate MMPs (10%)	800	7.63 (.279)	96.53 (3.521)	4.50 (.239)	3.51 (.229)	.26 (.181) ^{NS}	2.09 (.178)	5.52 (.170)
Many MMPs (20%)	800	21.07 (.477)	85.38 (1.845)	14.58 (.449)	10.94 (.371)	1.84 (.278)	6.96 (.277)	15.08 (.310)

Notes: Standard error of the mean is in parentheses. Differences are statistically significant at 99.9% level (paired sample t-test) unless marked with “NS.”

2018). Representing relationships in graphs supports two critical human capabilities: relational reasoning (i.e., drawing logical conclusions of how different objects relate to one another) and combinatorial generalization (i.e., constructing new inferences, predictions, and behaviors from known building blocks; Battaglia et al. 2018). The visualization of relationships among objects such as brands, products, or consumers thereby enables managers to solve complex problems that require the synthesis and analysis of diverse and extensive data (Sloman 1996).

The M4 framework proposed in this research is capable of visualizing objects in multiple locations on a single map. This capacity is crucial for obtaining accurate assessments of the structures of markets that contain products that compete in multiple submarkets. In a numerical example, an empirical application, and a simulation, this research shows that ignoring multimarket memberships of products, as alternative mapping methods do, can produce misleading representations of products' local competitive positions and may even distort the global visualization of the underlying market structure.

In this research, M4 is compared with alternative map projection methods of various research disciplines, including marketing (i.e., classic MDS, DRMAbs), scientometrics (i.e., VOS), and computer science (i.e., t-SNE). Although t-SNE is incorporated into the M4 framework in this research, alternative pairings are possible and are investigated in Web Appendix C. The analysis shows that, with the exception of classic MDS, all alternative mapping methods improve in the presence of MMPs when they are incorporated into the M4 framework.

Nevertheless, one should not conclude that MDS is generally inferior to alternative mapping methods solely because it performed poorly in the context of this research. MDS is an established method in marketing whose limitations are known. For instance, MDS solutions tend to deteriorate as the number of stimuli (e.g., products) increases—particularly for low-dimensional maps (Faure and Natter 2010). In addition, large stimuli sets may be sparse such that many relationships are missing or zero (see Ringel and Skiera 2016), which further degrades MDS's solution quality (Malhotra, Jain, and Pinson 1988). Indeed, in the case of the DSLR camera market study, over 30% of all possible relationships among DSLR cameras are never observed in over 57,000 clickstreams. However, in scenarios where all pairwise relationships among products are

available, their magnitudes are known, and triangular inequality is satisfied, classic MDS performs well—even in the presence of MMPs if MDS is integrated into the M4 framework (see Web Appendix H for a more elaborate discussion and empirical analysis).

Finally, MDS is a family of methods, of which classic MDS is used as a baseline in this research. Variants of MDS can provide additional benefits, such as modeling competitive asymmetry (e.g., Kim, Albuquerque, and Bronnenberg 2011), performing segmentation and positioning simultaneously (e.g., DeSarbo, Grewal, and Scott 2008), or explicitly considering consumer heterogeneity (e.g., Carroll and Chang 1970). To this end, a potential alternative to M4's MMT process might be INDSCAL (individual difference scaling analysis), which is an extension of classic MDS that considers consumer heterogeneity. To implement INDSCAL in the context of this article, submarkets would need to be conceptualized as layers in the third dimension of the three-way similarity matrix (i.e., submarket \times product \times product) that INDSCAL operates on. This idea is tested in Web Appendix I under the presence of MMPs. Although INDSCAL shows improvement over classic MDS across all performance metrics, it still falls short of what can be achieved with alternative mapping models.

That said, the M4 framework proposed herein is an addition to the literature on market segmentation and market structure visualization, not a replacement for previous methods. Methodologically, it overcomes a central constraint of alternative mapping methods: their requirement for objects (e.g., products) to appear in single map locations. By connecting overlapping clustering and map projection in a novel way, the M4 framework allows each product to appear in the local competitive context of every submarket it competes in. Compromises such as visualizing products somewhere between the submarkets they compete in are no longer necessary. Note that the idea of the MMT process, which is at the core of the M4 framework, could be operationalized in alternative ways to account for extremely large product assortments, for example (see Web Appendix J for a suggestion).

Substantively, this research shows that firms and researchers should reconsider the use of extant mapping methods in markets where multimarket memberships of products cannot be ruled out a priori. In case of the DSLR camera market, 50% of all submarkets were affected by MMPs. Unlike

traditional market structure maps, M4's maps enable managers to identify and investigate products' multimarket memberships within their local competitive contexts while preserving the market's overall (i.e., global) structure. As the empirical study demonstrates, a deeper investigation of phenomena observed in the M4 map can reveal how submarkets are affected by MMPs and whether the identified multimarket memberships are harmful or represent opportunities for firms. Managers can then use these insights to review and adjust their marketing mix decisions (e.g., adjust prices, communicate different positionings, change feature levels). By repeating the analysis over time, managers can track the impact of such adjustments and fine-tune their marketing mix further (Matthe, Ringel, and Skiera 2022). Researchers, in turn, can use M4 to quickly determine whether and where additional statistical analyses are required, what to look for, and where to look for it in the data. A browser-based tool, mapXP (www.mapxp.app), facilitates the interactive viewing, exploring, and sharing of market structure maps within firms and among researchers.

While this research demonstrates the value of M4, it is also subject to some limitations. First, as outlined at the beginning of this article, visual data discovery should not be understood as a substitute for deeper statistical analysis. Instead, the two are complements, whereby the former serves as a crucial starting point that informs the latter (Anscombe 1973). Second, the M4 framework is not appropriate for applications in which multimarket membership objects serve as transition points between clusters across which information, people, money, or other units flow. It may be more appropriate to visualize objects between clusters in such specific cases. An alternative possibility is to extend M4's map to incorporate arcs with arrowheads that connect objects to multimarket membership objects, thereby making the flows across clusters apparent (this functionality is already built into mapXP for future research applications). Third, M4 is not recommended for use in an empirical setting where submarkets are poorly separated (i.e., modularity < .20). Fourth, the sample of over 57,000 consumers searching for DSLR cameras at the price comparison site may not fully represent the entire population. Although this sample is substantial, a possible self-selection bias cannot be ruled out: some consumers may simply not search online, just as some consumers simply do not participate in surveys. Fifth, the M4 framework should not be perceived as a fully automated process that requires no judgment on the part of the analyst. In particular, it is recommended that analysts use those overlapping clustering and map projection methods within the M4 framework that are most appropriate for the particular empirical setting they are analyzing.

Nonetheless, it cannot be ruled out a priori that products meet consumer preferences of different submarkets and even compete across traditional product-market boundaries—particularly under high levels of product variety, complexity, and versatility (Day, Shocker, and Srivastava 1979; Shocker, Bayus, and Kim 2004; Urban, Johnson, and Hauser 1984; Yang, Zhang, and Kannan 2021). Instead of ignoring the complexity that arises from such relationship structures, the M4 framework embraces it and offers a first avenue toward better harnessing such complex

relationship structures in market structure visualization. Because the M4 framework is flexible and can easily be generalized to any system of relationships among objects, it is broadly applicable to various research problems in marketing and beyond.

Future directions for research include visualizing brand positionings and brand associations within and across multiple communities in social networks, investigating the effect of multimarket memberships on firm valuation and performance, identifying cobranding and copromotion opportunities in retail assortments on the basis of multicategory memberships of SKUs, discovering multitopic memberships in citation networks and their relationship to research impact, studying consumer needs and choices (that depend on usage context and situation), and tracing supply chains (where firms are associated at various levels), to name a few. On the methodological front, potential extensions of the M4 framework include explicitly modeling consumer heterogeneity, explicitly modeling brand product hierarchies, capturing competitive asymmetry, and analyzing the formation and evolution of MMPs' positions over time.

Appendix

The MMT process can be described formally as follows: Let J be an index set of all products in a market under consideration, $J = \{1, \dots, |J|\}$, whose competitive relationships are captured in a $|J| \times |J|$ relationship matrix S . The j th row and j th column of matrix S are denoted by S_j and $S_{\cdot j}$ respectively. Each cell S_{ij} , $j \in J$, captures the competitive relationship between two products. Further, let C be the index set of all submarkets, with submarket $c \in \{1, \dots, C\}$. Let O_c be the set of products from J that compete in submarket c . Finally, let \mathbf{a} be a vector that holds the submarkets $c \in C$ that product j is assigned to by SpeakEasy, with $\mathbf{a}_{j,i} \in \{1, \dots, M\}$ holding the i th submarket assignment of product j , and M being the number of submarkets product j is assigned to.

Multiple instances of each product j that was assigned to multiple submarkets by SpeakEasy (where $M > 1$) are created in matrix S by duplicating j 's respective rows and columns to achieve M instances for j in total. All irrelevant relationships are then removed from each instance of j so that each instance pertains to the specific competitive context of its submarket. Specifically, a loop over Equations A1 through A10 is performed for each MMP j (for $M > 1$) and for all $\mathbf{a}_{j,i>1}$ (that is, for all submarket assignments of j except the first one). The first step is to construct the union set of all products that are assigned to submarkets that MMP j is also assigned to, except for those products that are assigned to submarket $\mathbf{a}_{j,i}$:

$$\bar{O}_{c=\mathbf{a}_{j,i}} = \left\{ x \in \bigcup_{k=1}^M O_{c=\mathbf{a}_{j,k}} \mid x \notin O_{c=\mathbf{a}_{j,i}} \right\}. \quad (\text{A1})$$

The set $\bar{O}_{c=\mathbf{a}_{j,i}}$ now contains those products of J whose competitive relationships to the instance of MMP j assigned to submarket $\mathbf{a}_{j,i}$ must be removed. Next, the set of "global competitors" of MMP j (i.e., products that have competitive relationships

with MMP j but that do not compete in any of the submarkets that MMP j competes in) is identified as follows:

$$G = \left\{ x \in J \mid x \notin \bigcup_{k=1}^M O_{c=a_{j,k}} \right\}. \quad (A2)$$

To assign global competitive relationships, the local representative(s) of each instance of MMP j are first identified. Note that an instance can have multiple local representatives when it has identical “strong” relationships with multiple products within the focal submarket. For each element of vector \mathbf{a} (that is, for each instance of MMP j), the set of products $R_{t \in \{1, \dots, M\}}$ from J assigned to submarket $c = \mathbf{a}_{j,t}$ that have the strongest competitive relationships to MMP j is identified as follows:

$$R_t = \left\{ x \in O_{c=a_{j,t}} \mid x \in \arg \max_x f(S_{jx}) \right\}. \quad (A3)$$

Using the representatives R_t for all instances of MMP j , the next step is to identify the set of products \bar{G}_i from which the competitive relationships to the instance of MMP j in submarket $\mathbf{a}_{j,i}$ must be removed:

$$\bar{G}_i = \left\{ x \in G \mid \left(\sum_{k \in R_{t=i}} S_{kx} \right) / |R_{t=i}| < \max_{t=1}^M \left(\sum_{k \in R_t} S_{kx} \right) / |R_t| \right\}. \quad (A4)$$

Since a single instance of MMP j can have multiple representatives, their mean global competitive relationship weight to any given global product is used to determine which global competitive relationships to remove. Next, $\bar{O}_{c=a_{j,i}}$ and \bar{G}_i are used to construct a $1 \times |J|$ matrix \mathbf{Z} that identifies the products in J for which the competitive relationships of the duplicated instance in submarket $\mathbf{a}_{j,i}$ of MMP j must be assigned (i.e., set to 1) and removed (i.e., set to 0):

$$\mathbf{Z}_k = k \in \left\{ \bar{O}_{c=a_{j,i}} \bigcup \bar{G}_i \right\} \rightarrow 0, 1. \quad (A5)$$

Then, a new relationship matrix $\tilde{\mathbf{S}}$ that includes the newly created instance of MMP j for submarket $\mathbf{a}_{j,i}$ is constructed by appending the Hadamard product of \mathbf{S}_j and \mathbf{Z} (as a row) as well as of \mathbf{S}_j and \mathbf{Z}^T (as a column) to matrix \mathbf{S} :

$$\tilde{\mathbf{S}} = \begin{bmatrix} \mathbf{S} & \mathbf{S}_j \circ \mathbf{Z}^T \\ \mathbf{S}_j \circ \mathbf{Z} & \mathbf{S}_{jj} \end{bmatrix}. \quad (A6)$$

To include the new instance of MMP j in matrix $\tilde{\mathbf{S}}$ in the analysis, the index set J is updated to contain $|J| + 1$ products:

$$J = J \cup \{|J| + 1\}. \quad (A7)$$

The next step is to identify which competitive relationships to MMP j 's instance in submarket \mathbf{a}_1 (that is, the row and column of original product j) must be removed from matrix $\tilde{\mathbf{S}}$ so that they are not duplicated across its instances. These competitive relationships include local competitive relationships to products in submarket $\mathbf{a}_{j,i}$ of set $O_{c=a_{j,i}}$ and global competitive relationships to products that have stronger competitive relationships to the representatives of other instances of MMP j (that is, $R_{t \neq i}$). The latter ($\bar{G}_{i=1}$) is identified by setting $i = 1$ in Equation A4. Then, the j th row and j th column of matrix $\tilde{\mathbf{S}}$ are

replaced with the respective Hadamard products of $\tilde{\mathbf{S}}_j$ and $\tilde{\mathbf{Z}}$ ($\tilde{\mathbf{S}}_j$ and $\tilde{\mathbf{Z}}^T$):

$$\tilde{\mathbf{S}}_j = \tilde{\mathbf{S}}_j \circ \tilde{\mathbf{Z}}, \quad (A8)$$

$$\tilde{\mathbf{S}}_j = \tilde{\mathbf{S}}_j \circ \tilde{\mathbf{Z}}^T, \quad (A9)$$

where $\tilde{\mathbf{Z}}$ is a $1 \times |J|$ matrix that identifies the products in J for which the competitive relationships of the instance of MMP j in submarket $\mathbf{a}_{j,1}$ (that is, the original product j) must be assigned (set to 1) and removed (set to 0):

$$\tilde{\mathbf{Z}}_k = k \in \left\{ O_{c=a_j} \bigcup \bar{G}_1 \right\} \rightarrow 0, 1. \quad (A10)$$

Finally, matrix \mathbf{S} is replaced with matrix $\tilde{\mathbf{S}}$ ($\mathbf{S} \leftarrow \tilde{\mathbf{S}}$) so that the next iteration of the loop is based on matrix $\tilde{\mathbf{S}}$. After the final loop, all MMPs in J have local instances in matrix $\tilde{\mathbf{S}}$ that pertain to the specific competitive context of their submarkets. Various map projection methods can now be used to visualize relationship matrix \mathbf{S} .

Acknowledgments

This article is based on the author's dissertation. The author thanks an anonymous price-comparison site for supplying search data for this research. The author is particularly grateful for the support and valuable feedback of Bernd Skiera, Jan-Benedict Steenkamp, and Valarie Zeithaml. The author thanks the anonymous *JMR* review team for their many constructive comments and excellent guidance. This article also profited from the comments and suggestions of André Martin and Maximilian Matthe, as well as from the feedback of the marketing faculties at Dartmouth College, George Washington University, Goethe University Frankfurt, the University of North Carolina at Chapel Hill, and the University of Technology Sydney.

Associate Editor

Eric Bradlow


Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the German Research Foundation, Deutsche Forschungsgemeinschaft (grant number SK 66/9-1).

ORCID iD

Daniel M. Ringel  <https://orcid.org/0000-0002-2070-000X>

References

- Akkucuk, Ulas and J. Douglas Carroll (2006), “Paramap vs. Isomap: A Comparison of Two Nonlinear Mapping Algorithms,” *Journal of Classification*, 23 (2), 221–54.
- Andrews, Rick L. and Ajay K. Manrai (1999), “MDS Maps for Product Attributes and Market Response: An Application to Scanner Panel Data,” *Marketing Science*, 18 (4), 584–604.

- Anscombe, Francis J. (1973), "Graphs in Statistical Analysis," *The American Statistician*, 27 (1), 17–21.
- Antoniak, Maria and David Mimno (2018), "Evaluating the Stability of Embedding-Based Word Similarities," *Transactions of the Association for Computational Linguistics*, 6, 107–19.
- Arabie, Phipps, J. Douglas Carroll, Wayne S. DeSarbo, and Jerry Wind (1981), "Overlapping Clustering: A New Method for Product Positioning," *Journal of Marketing Research*, 18 (3), 310–17.
- Battaglia, Peter W., Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, et al. (2018), "Relational Inductive Biases, Deep Learning, and Graph Networks," preprint, arXiv, <https://doi.org/10.48550/arXiv.1806.01261>.
- Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre (2008), "Fast Unfolding of Communities in Large Networks," *Journal of Statistical Mechanics: Theory and Experiment*, 2008 (10), P10008.
- Branco, Fernando, Monic Sun, and J. Miguel Villas-Boas (2012), "Optimal Search for Product Information," *Management Science*, 58 (11), 2037–56.
- Bronnenberg, Bart J., Byeong Jun Kim, and Carl F. Mela (2016), "Zooming In on Choice: How Do Consumers Search for Cameras Online?" *Marketing Science*, 35 (5), 693–712.
- Carroll, J. Douglas and Jih-Jie Chang (1970), "Analysis of Individual Differences in Multidimensional Scaling via an N-Way Generalization of 'Eckart-Young' Decomposition," *Psychometrika*, 35 (3), 283–319.
- Chaturvedi, Anil, J. Douglas Carroll, Paul E. Green, and John A. Rotondo (1997), "A Feature-Based Approach to Market Segmentation Via Overlapping K-Centroids Clustering," *Journal of Marketing Research*, 34 (3), 370–77.
- Collins, Linda M. and Clyde W. Dent (1988), "Omega: A General Formulation of the Rand Index of Cluster Recovery Suitable for Non-Disjoint Solutions," *Multivariate Behavioral Research*, 23 (2), 231–42.
- Day, George S., Allan D. Shocker, and Rajendra K. Srivastava (1979), "Customer-Oriented Approaches to Identifying Product-Markets," *Journal of Marketing*, 43 (4), 8–19.
- De Meo, Pasquale, Emilio Ferrara, Giacomo Fiumara, and Alessandro Provetti (2014), "Mixing Local and Global Information for Community Detection in Large Networks," *Journal of Computer and System Sciences*, 80 (1), 72–87.
- DeSarbo, Wayne S., Rajdeep Grewal, and Crystal J. Scott (2008), "A Clusterwise Bilinear Multidimensional Scaling Methodology for Simultaneous Segmentation and Positioning Analyses," *Journal of Marketing Research*, 45 (3), 280–92.
- DeSarbo, Wayne S., Ajay K. Manrai, and Lalita A. Manrai (1993), "Non-Spatial Tree Models for the Assessment of Competitive Market Structure: An Integrated Review of the Marketing and Psychometric Literature," *Handbooks in Operations Research and Management Science*, 5, 193–257.
- Elrod, Terry (1988), "Choice Map: Inferring a Product-Market Map from Panel Data," *Marketing Science*, 7 (1), 21–40.
- Faure, Corinne and Martin Natter (2010), "New Metrics for Evaluating Preference Maps," *International Journal of Research in Marketing*, 27 (3), 261–70.
- Fershtman, Daniel and Alessandro Pavan (2020), "Sequential Learning with Endogenous Consideration Sets," CEPR Discussion Paper No. DP15018, <https://ssrn.com/abstract=3650133>.
- France, Stephen L. and Sanjoy Ghose (2016), "An Analysis and Visualization Methodology for Identifying and Testing Market Structure," *Marketing Science*, 35 (1), 182–97.
- France, Stephen L. and Sanjoy Ghose (2018), "Marketing Analytics: Methods, Practice, Implementation, and Links to Other Fields," *Expert Systems with Applications*, 119, 456–75.
- Franke, Nikolaus, Heribert Reisinger, and Daniel Hoppe (2009), "Remaining Within-Cluster Heterogeneity: A Meta-Analysis of the Dark Side of Clustering Methods," *Journal of Marketing Management*, 25 (3/4), 273–93.
- Gabel, Sebastian, Daniel Guhl, and Daniel Klapper (2019), "P2V-MAP: Mapping Market Structures for Large Retail Assortments," *Journal of Marketing Research*, 56 (4), 557–80.
- Gaiteri, Chris, Mingming Chen, Boleslaw Szymanski, Konstantin Kuzmin, Jierui Xie, Changkyu Lee, et al. (2015), "Identifying Robust Communities and Multi-Community Nodes by Combining Top-Down and Bottom-Up Approaches to Clustering," *Scientific Reports*, 5, 16361.
- Gardete, Pedro and Megan Hunter (2020), "Guiding Consumers Through Lemons and Peaches: An Analysis of the Effects of Search Design Activities," Stanford University Graduate School of Business Research Paper No. 3669, <https://ssrn.com/abstract=3180811>.
- Gartner (2018), *Magic Quadrant for Analytics and Business Intelligence Platforms*, Cindi Howson, Rita L. Sallam, James Laurence Richardson, Joao Tapadinhas, Carlie J. Idoine, and Alys Woodward, eds. Gartner.
- Green, Paul E. (1975), "Marketing Applications of MDS: Assessment and Outlook," *Journal of Marketing*, 39 (1), 24–31.
- Green, Paul E., Yoram Wind, and Henry J. Claycamp (1975), "Brand-Features Congruence Mapping," *Journal of Marketing Research*, 12 (3), 306–13.
- Greminger, Rafael P. (2022), "Optimal Search and Discovery," *Management Science*, 68 (5), 3904–24.
- Grover, Rajiv and William R. Dillon (1985), "A Probabilistic Model for Testing Hypothesized Hierarchical Market Structures," *Marketing Science*, 4 (4), 312–35.
- Grover, Rajiv and V. Srinivasan (1987), "A Simultaneous Approach to Market Segmentation and Market Structuring," *Journal of Marketing Research*, 24 (2), 139–53.
- Hausmann, Jean-Claude (1995), "On the Vietoris-Rips Complexes and a Cohomology Theory for Metric Spaces," in *Annals of Mathematics Studies: Prospects in Topology*, Frank Quinn, ed. Princeton University Press, 175–88.
- Heer, Jeffrey and Ben Shneiderman (2012), "Interactive Dynamics for Visual Analysis," *Queue*, 10 (2), 30–55.
- Iacobucci, Dawn, Doug Grisaffe, and Wayne S. DeSarbo (2017), "Statistical Perceptual Maps: Using Confidence Region Ellipses to Enhance the Interpretations of Brand Positions in Multidimensional Scaling," *Journal of Marketing Analytics*, 5 (3/4), 81–98.
- Kannan, P.K. and Susan M. Sanchez (1994), "Competitive Market Structures: A Subset-Selection Analysis," *Management Science*, 40 (11), 1484–99.

- Kim, Jun B., Paulo Albuquerque, and Bart J. Bronnenberg (2011), "Mapping Online Consumer Search," *Journal of Marketing Research*, 48 (1), 13–27.
- Lancichinetti, Andrea and Santo Fortunato (2012), "Consensus Clustering in Complex Networks," *Scientific Reports*, 2 (1), 336.
- Lancichinetti, Andrea, Santo Fortunato, and Filippo Radicchi (2008), "Benchmark Graphs for Testing Community Detection Algorithms," *Physical Review E*, 78 (4), 046110.
- Lattin, James M. and Leigh McAlister (1985), "Using a Variety-Seeking Model to Identify Substitute and Complementary Relationships Among Competing Products," *Journal of Marketing Research*, 22 (3), 330–39.
- Lee, Thomas Y. and Eric T. Bradlow (2011), "Automated Marketing Research Using Online Customer Reviews," *Journal of Marketing Research*, 48 (5), 881–94.
- Lilien, Gary L. and Arvind Rangaswamy (2004), *Marketing Engineering: Computer-Assisted Marketing Analysis and Planning*. DecisionPro.
- Malhotra, Naresh K., Arun K. Jain, and Christian Pinson (1988), "The Robustness of MDS Configurations in the Case of Incomplete Data," *Journal of Marketing Research*, 25 (1), 95–102.
- Marbeau, Yves (1998), *Communication of Research Results*, 4th ed. ESOMAR.
- Matthe, Maximilian, Daniel M. Ringel, and Bernd Skiera (2022), "Mapping Market Structure Evolution," *Marketing Science* (published online September 8), <https://doi.org/10.1287/mksc.2022.1385>.
- McInnes, Leland and John Healy (2018), "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," preprint, arXiv, <https://doi.org/10.48550/arXiv.1802.03426>.
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean (2013), "Distributed Representations of Words and Phrases and Their Compositionality," *Advances in Neural Information Processing Systems*, 26, 3111–19.
- Moe, Wendy W. (2006), "An Empirical Two-Stage Choice Model with Varying Decision Rules Applied to Internet Clickstream Data," *Journal of Marketing Research*, 43 (4), 680–92.
- Nam, Hyoryung, Yogesh V. Joshi, and P.K. Kannan (2017), "Harvesting Brand Information from Social Tags," *Journal of Marketing*, 81 (4), 88–108.
- Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko (2012), "Mine Your Own Business: Market-Structure Surveillance Through Text Mining," *Marketing Science*, 31 (3), 521–43.
- Rao, Vithala R. and Ralph Katz (1971), "Alternative Multidimensional Scaling Methods for Large Stimulus Sets," *Journal of Marketing Research*, 8 (4), 488–94.
- Ringel, Daniel M. and Bernd Skiera (2016), "Visualizing Asymmetric Competition Among More Than 1,000 Products Using Big Search Data," *Marketing Science*, 35 (3), 511–34.
- Sammon, John W. (1969), "A Nonlinear Mapping for Data Structure Analysis," *IEEE Transactions on Computers*, 100 (5), 401–09.
- Shocker, Allan D., Barry L. Bayus, and Namwoon Kim (2004), "Product Complements and Substitutes in the Real World: The Relevance of 'Other Products,'" *Journal of Marketing*, 68 (1), 28–40.
- Shocker, Allan D., David W. Stewart, and Anthony J. Zahorik (1990), "Determining the Competitive Structure of Product-Markets: Practices, Issues, and Suggestions," *Journal of Managerial Issues*, 2 (2), 127–59.
- Sloman, Steven A. (1996), "The Empirical Case for Two Systems of Reasoning," *Psychological Bulletin*, 119 (1), 3–22.
- Smelcer, John B. and Erran Carmel (1997), "The Effectiveness of Different Representations for Managerial Problem Solving: Comparing Tables and Maps," *Decision Sciences*, 28 (2), 391–420.
- Sobolevsky, Stanislav, Riccardo Campari, Alexander Belyi, and Carlo Ratti (2014), "General Optimization Technique for High-Quality Community Detection in Complex Networks," *Physical Review E*, 90 (1), 012811.
- Thorelli, Hans B. (1986), "Networks: Between Markets and Hierarchies," *Strategic Management Journal*, 7 (1), 37–51.
- Tobler, Waldo R. (1970), "A Computer Movie Simulating Urban Growth in the Detroit Region," *Economic Geography*, 46 (Suppl. 1), 234–40.
- Trope, Yaacov and Nira Liberman (2010), "Construal-Level Theory of Psychological Distance," *Psychological Review*, 117 (2), 440–63.
- Urban, Glen L., Philip L. Johnson, and John R. Hauser (1984), "Testing Competitive Market Structures," *Marketing Science*, 3 (2), 83–112.
- Ursu, Raluca M., Qingliang Wang, and Pradeep K. Chintagunta (2020), "Search Duration," *Marketing Science*, 39 (5), 849–71.
- Van der Maaten, Laurens and Geoffrey Hinton (2008), "Visualizing Data Using t-SNE," *Journal of Machine Learning Research*, 9 (November), 2579–2605.
- Van Eck, Nees J., Ludo Waltman, Rommert Dekker, and Jan van den Berg (2010), "A Comparison of Two Techniques for Bibliometric Mapping: Multidimensional Scaling and VOS," *Journal of the American Society for Information Science and Technology*, 61 (12), 2405–16.
- Van Heerde, Harald J., Shuba Srinivasan, and Marnik G. Dekimpe (2010), "Estimating Cannibalization Rates for Pioneering Innovations," *Marketing Science*, 29 (6), 1024–39.
- Voleti, Sudhir, Praveen K. Kopalle, and Pulak Ghosh (2015), "An Interproduct Competition Model Incorporating Branding Hierarchy and Product Similarities Using Store-Level Data," *Management Science*, 61 (11), 2720–38.
- Wedel, Michel and Jan-Benedict E.M. Steenkamp (1991), "A Clusterwise Regression Method for Simultaneous Fuzzy Market Structuring and Benefit Segmentation," *Journal of Marketing Research*, 28 (4), 385–96.
- Yang, Yi, Kunpeng Zhang, and P.K. Kannan (2021), "Identifying Market Structure: A Deep Network Representation Learning of Social Engagement," *Journal of Marketing*, 86 (4), 37–56.
- Zahorik, Anthony J. (1994), "A Nonhierarchical Brand Switching Model for Inferring Market Structure," *European Journal of Operational Research*, 76 (2), 344–58.