Article

Journal of Marketing Research
2019, Vol. 56(4) 557-580

DOI: 10.1177/0022243719833631
journals.sagepub.com/home/mrj

P2V-MAP: Mapping Market Structures for Large Retail Assortments

Sebastian Gabel, Daniel Guhl, and Daniel Klapper

Abstract

The authors propose a new, exploratory approach for analyzing market structures that leverages two recent methodological advances in natural language processing and machine learning. They customize a neural network language model to derive latent product attributes by analyzing the co-occurrences of products in shopping baskets. Applying dimensionality reduction to the latent attributes yields a two-dimensional product map. This method is well-suited to retailers because it relies on data that are readily available from their checkout systems and facilitates their analyses of cross-category product complementarity, in addition to within-category substitution. The approach has high usability because it is automated, is scalable and does not require a priori assumptions. Its results are easy to interpret and update as new market basket data are collected. The authors validate their approach both by conducting an extensive simulation study and by comparing their results with those of state-of-the-art, econometric methods for modeling product relationships. The application of this approach using data collected at a leading German grocery retailer underlines its usefulness and provides novel findings that are relevant to assortment-related decisions.

Keywords

market structure analysis, product maps, machine learning, retailing, big data

Online supplement: https://doi.org/10.1177/0022243719833631

The analysis of market structures that reflect substitution and complementarity between products and services is fundamental to marketing research and practice (Elrod et al. 2002), especially in retail settings. Online and offline retailers must regularly make decisions that depend directly on their understanding of product relationships (Mantrala et al. 2009). Firms must choose how many distinct categories to carry and how many products to offer in each category (Kök, Fisher, and Vaidyanathan 2015). Brick-and-mortar retailers want to optimize product arrangements in aisles and shelves; online retailers strive to identify the best way to sort and group products in their online shops (Breugelmans, Campo, and Gijsbrechts 2007; Van Nierop, Fok, and Franses 2008). Furthermore, retailers must decide which items to promote simultaneously, such as in leaflets or newsletters. These assortment-related decisions influence consumer store choices, sales, and profits (Briesch, Chintagunta, and Fox 2009; Kök, Fisher, and Vaidyanathan 2015; Mantrala et al. 2009; Stassen, Mittelstaedt, and Mittelstaedt 1999; Venkatesan and Farris 2012).

Despite the significance of market structures for retailers, many studies adopt a manufacturer's perspective and analyze the competition among a small set of previously selected products, typically within a single product category. Restricting the analysis in this way excludes the majority of a retailer's product portfolio. Moreover, the definition of product categories ex ante entails strong assumptions about product substitution and neglects product relationships across categories. In addition, a category-centric view often produces little insight into market structures. For example, consumers rarely purchase several units of wine in the same shopping trip. In cases such as this one, market basket data make it possible to learn about market structures and product attributes, not from the product purchase itself but from the product's basket co-occurrence with other products.

Elrod et al. (2002) argue that market structure analyses tend to focus on substitution in narrowly defined product categories because of the limited availability of cheap data, computing power, or scalable modeling technologies. Today, retailers benefit from substantial improvements in computing power and

Sebastian Gabel is Chief AI Officer, SO1 GmbH, and a doctoral student, School of Business and Economics, Humboldt University Berlin, Germany (email: sebastian.gabel@hu-berlin.de). Daniel Guhl is a Postdoctoral Researcher, Humboldt University Berlin, Germany (email: daniel.guhl@hu-berlin.de). Daniel Klapper is Professor of Marketing, Humboldt University Berlin, Germany (email: daniel.klapper@hu-berlin.de).

can collect overwhelming amounts of high-quality data (Bradlow et al. 2017). Nonetheless, to the best of our knowledge, no marketing literature has investigated market structures across all products in a retailer's assortment, mainly because three shortcomings of existing methods limit their application. First, these approaches cannot scale adequately to big data and large product portfolios. Consider, for example, that Walmart collects data on billions of shopping baskets every year and stocks up to 150,000 distinct products in its brick-and-mortar stores, along with more than 1 million products on walmart.com (Walmart 2005, 2016). Second, these approaches are not capable of modeling both drivers of market structures that are reflected by market basket data: in addition to within-category competition, product complementarity across categories merits investigation (Elrod et al. 2002). Third, visualization techniques often produce maps that lump products together, making the results difficult to interpret (Netzer et al. 2012).

With this article, we aim to (1) develop P2V-MAP (short for product2vec-map), a new approach for analyzing market structures that meets the needs of retailers, (2) validate our approach with simulated data, (3) apply and validate it in an empirical setting, and (4) provide suggestions on how the insights about the market structure generated by P2V-MAP can inform assortment-related decisions. P2V-MAP is based on two recent methodological advances in natural language processing (NLP) and machine learning: the skip-gram (SG) model, a neural network language model (Mikolov et al. 2013), and the Barnes–Hut t-distributed stochastic neighbor embedding (t-SNE) dimensionality reduction algorithm (Van der Maaten 2014). First, we extend the SG model to allow for modeling market basket data and then use it to derive latent attributes for all products in a retailer's product assortment; that is, we model co-occurrences of products in shopping baskets. Second, we create a two-dimensional map of the complete product assortment by applying t-SNE to the SG model's output. The map is based on latent product attributes derived from the co-occurrence of products in shopping baskets, so it provides insights into the consumers' view of the market structure and basket composition.

Our work is relevant to both research and practice, and it extends existing methods in three ways. First, P2V-MAP is retailer-centric, considering a retailer's whole product assortment and analyzing both within-category product substitution and cross-category product complementarity (Elrod et al. 2002). Second, P2V-MAP offers high usability. Specifically, the creation of the product map is fully automated and does not require human interaction, such as to define product categories. It leverages data that are already available to retailers directly from their checkout systems, so it is both cost effective and universal. Because it is based on scalable machine learning methods, it can be applied to billions of shopping baskets and vast assortments, and results can be updated incrementally with new market basket data. Finally, t-SNE produces product maps that do not lump products together, as existing approaches often do. Product clusters are well separated because the latent attributes capture the attributes that differentiate products well. Third, we present relevant insights into market structures. For the grocery retailer analyzed in this study, we find many product clusters that contain products from multiple categories; the common denominators in these clusters are specific product attributes (e.g., organic, gluten free, meant for children). The product map also provides strong evidence of product complementarity across retailer categories (e.g., hot dog/burger ingredients). Within product clusters, the map indicates how consumers differentiate products. In the wine cluster, for example, price is the attribute that offers the strongest differentiation, although the retailer sorts wine shelves by color and country of origin. For chocolate bars, we find that the retailer carries various products that consumers regard as almost identical. Our findings have face validity and generate insights that go beyond current views of market structures. We verify each result using comparisons with well-established econometric methods for modeling product relationships; in contrast to econometric approaches, P2V-MAP produces all results in a single analysis. These results suggest that P2V-MAP can support practitioners in aisle and shelf design, assortment decisions, and promotion management. A sample application illustrates how the market clusters derived from P2V-MAP enable retailers to analyze their consumers' store-specific preferences and how these preferences relate to the assortment depth across clusters.

The remainder of this article is organized as follows: After reviewing the relevant literature, we formally introduce P2V-MAP. Then, we validate the proposed method through an extensive simulation study and apply it to a large basket data set from a leading German grocery retailer. We present the findings and validate them through comparisons with the results of state-of-the-art methods for modeling product relationships. After summarizing our main results and discussing the managerial implications, we conclude by providing directions for further research.

Relevant Literature

Three streams of literature are highly relevant to our work: research on product maps, market basket analysis, and machine learning methods. We add to research on product maps (for an overview, see Ringel and Skiera [2016]) in four ways. First, we map products according to their co-occurrence patterns in a large number of shopping baskets. Doing so simultaneously for all products in a retailer's assortment facilitates the analysis of product complementarity, whereas existing methods for creating product maps focus only on competitive relationships. Creating one map across all product categories in a single analysis also eliminates the need to make a priori assumptions about substitution by selecting the products to include in the analysis. Extracting deep insights from raw market basket data makes P2V-MAP very suitable for dashboards, which are the primary basis for decision making in industry (Wedel and Kannan 2016).

Second, our approach based on shopping baskets invokes a cheap data source that is available to all retailers through their checkout systems. It does not require customer reviews (Lee and Bradlow 2011), search data (Kim, Albuquerque, and Bronnenberg 2011; Ringel and Skiera 2016), online discussion forums (Netzer et al. 2012), survey data (DeSarbo and Grewal 2007), or data that track individual consumers' purchases over time (Erdem 1996). Instead of relying on external information from secondary data sources, our method derives insights about the market structure solely from basket data such that it can analyze market structures in situations in which other data sources might be expensive or difficult to acquire. Basket data are a reliable and valid source to reveal consumers' views of products and market structures.

Third, the method is very scalable. The SG model can be estimated in batches, which enables us to tap into billions of shopping baskets to derive latent product attributes. The low training complexity of the SG model, which is proportional to the logarithm of the number of products (Mikolov et al. 2013), and the low computational complexity of t-SNE (Van der Maaten 2014) make it possible to apply P2V-MAP to map the market structures of online retailers, which often carry millions of products in their portfolios.

Fourth, using latent attributes enables us to automate the analysis of market structures. Typically, "regardless of an internal or external approach, with few exceptions, market structure analysis begins with a predetermined set of attributes" (Lee and Bradlow 2011, p. 882). Our approach extends work by Lee and Bradlow (2011), who use text mining to generate and select attributes automatically from product reviews for durable goods. Avoiding product reviews is practical when it comes to low-involvement products; the useful attributes likely cannot be derived if products seem less important and the decision processes are less involved. Our proposed method is universal, in that it can map any type of product for which checkout data are available without requiring human interaction. Using text data to mine product attributes also tends to yield distinct attributes per category. The common latent attribute space for all products simplifies the simultaneous mapping of products, because it makes them comparable, even if they represent different categories. This trait is key for creating a single market structure map for a retailer's entire assortment and therefore analyzing product complementarity. In addition, P2V-MAP eliminates the need to make a priori assumptions to define or select the product attributes, which in turn minimizes the risk of omitting attributes that might be important to consumers or developing biased views of the market structure.

In regard to market basket analysis, it is important to note that previous research has analyzed product complementarity before. Multivariate probit models (e.g., Manchanda, Ansari, and Gupta 1999) and multivariate logit models (e.g., Russell and Petersen 2000) replace a consumer's choice of multiple items with a single choice of an item bundle. Although such models do not depict market structures, they provide cross-effects that reveal product complementarity across categories. In the case of large assortments, however, the estimation of these models becomes computationally intractable (Hruschka 2014). Furthermore, researchers need to select a limited set of categories to be modeled a priori. In contrast, P2V-MAP can be applied simultaneously to all products and categories in a retailer's assortment without requiring assumptions about the categories to which the products belong. In this sense, our approach complements a market basket analysis because researchers can use it as a tool for category delineation, essentially supporting the educated delineation and selection of categories for analysis in multicategory models.

Recent applications of latent Dirichlet allocation (LDA) to market basket analysis (Jacobs, Donkers, and Fok 2016) also have uncovered associations between observed purchases and the underlying preferences of consumers, yielding so-called topic vectors for products and baskets. Similar to the latent attribute vectors inferred by our approach, these topic vectors can be used for visualization. The difference is that P2V-MAP vectors are dense, whereas LDA vectors are sparse. Sparser topic vectors simplify the interpretation of the results, readily revealing which products belong to the same topic, but denser latent attribute vectors yield more flexible similarity patterns, especially if the products belong to different market clusters. Another important feature of the dense P2V-MAP vectors is their linearity. This makes it possible to infer the vector of a product through a linear combination of other product vectors. In this article, we show that both aspects are useful for mapping market structures. In addition, LDA becomes computationally expensive for larger data sets (Mikolov et al. 2013), so applications in marketing tend to use rather small data sets. For example, Jacobs, Donkers, and Fok (2016) use 95,208 purchases of 394 products. P2V-MAP can derive market structures from billions of shopping baskets and works well for large product assortments.

With these elements, this study also contributes to methodological research pertaining to machine learning and marketing. Most importantly, we show how to modify the SG model to the application to market basket data. Our work also illustrates new applications of the SG model and t-SNE. Beyond NLP, these methods have been applied to gene sequencing (Asgari and Mofrad 2015) and online recommender systems (Phi, Chen, and Hirate 2016); we provide the first application to market basket data, thereby extending the models' applicability. Furthermore, transferring machine learning models to marketing applications is relevant to marketing researchers. When dealing with large data sets, it becomes increasingly important "to understand fundamental research in areas such as machine learning and natural language processing ... and apply these appropriately in marketing contexts" (Sudhir 2016, p. 4). We further such developments by depicting precisely how to apply the SG model and t-SNE to marketing questions. Deriving representations of products in latent attribute spaces also lays the foundation for deep learning applications in marketing; the latent attributes that result from the SG model can be used as input layers for deep neural networks (LeCun, Bengio, and Hinton 2015).

A Machine Learning Approach to Market Structure Analysis

P2V-MAP is a scalable approach for the visualization and analysis of market structures. It is based on two algorithms from NLP and machine learning: the SG model (Mikolov et al. 2013) and t-SNE (Van der Maaten 2014). So far, the SG model has not been applied to market basket data. Therefore, we explain how to modify the SG model to model the co-occurrence of products in shopping baskets, how to prepare basket data as input for the SG model, how to use the model's output as input for t-SNE, and how to enrich the output of t-SNE with additional information about the market structure. We discuss why the two algorithms are suitable for mapping market structure and propose suitable values for hyperparameters. In total, our approach consists of the three steps in Figure 1: (1) data preparation, (2) latent product attributes (SG model), and (3) mapping of products (t-SNE) and attribute overlays (e.g., visualizing attributes by varying the colors and sizes of bubbles that represent products). We refer to a product's vector of latent attributes as a product vector or vector representation and to the matrix of latent attributes over all products as a product embedding.

Step I: Data Preparation

The input data for the SG model are available from the retailer's checkout system. A shopping basket b, identified by a unique basket identifier (ID), is a set of n_b products purchased together in a single transaction by the same consumer at the same time in the same store. A product j in product assortment J is identified by its global trade item number. The only input data for the SG model are a map between basket and product IDs $b_i \rightarrow \{ j_1^{b_i}, \ldots, j_{n_b}^{b_i} \}$; no master data (e.g., product descriptions, category labels) are required to derive latent attributes.

To use basket data in the next stages of our approach, we must undertake three data preprocessing steps. First, certain products are rarely observed in basket data. During the training, product vectors for such products essentially remain in their random initialization state and only add noise to the model; therefore, we remove products that are observed in fewer than n_{min} shopping baskets. Mikolov et al. (2013) also propose subsampling frequent products. In NLP, the most frequent words, such as "a" and "the," occur in a large fraction of sentences and typically provide less information than less frequent words. In the basket data used in our empirical application, the most frequent product occurs in only 4.1% of all shopping baskets; therefore, the subsampling of product occurrences is not necessary and would remove valuable information from the data set.

Second, we build the actual training samples for the SG model by reshaping the basket data (see Figure 1). We follow the notation of NLP and refer to the training samples as center-context pairs. For each product in a shopping basket (center product), we create all combinations of the product with the remaining products in the shopping basket (context products), yielding $n_b(n_b - 1)$ positive training samples for each shopping basket b. Note that this approach for generating training samples differs from the original SG model (Mikolov et al. 2013). In NLP, training samples are only constructed from words within a context window $\pm w$ around a center word. Words that are closer to each other in a sentence are often related, so focusing on a limited window around a word is reasonable for language modeling. However, this is not the case in our application, in which the order of products in shopping baskets is random. We can maximize the amount of information used to infer product vectors by using all products in a basket, not only those from an artificially constructed window.

Third, we create negative samples, that is, product pairs that are not observed together in the shopping basket. Product vectors are trained such that the SG model can accurately classify whether a center-context pair is a positive or a negative sample. For each positive training example, we sample n_{neg} products from total product assortment J using a probability distribution with density $P_n(j) = (1/Z)\, n_j^{pow}$ as sampling weights; n_j is the number of times product j occurs in training data, Z is a normalization constant, and pow is a constant that allows for modifying the shape of the sampling distribution. Common values for n_{neg} are 2 and 20, and pow is typically set to .75 (Mikolov et al. 2013). In contrast to the original SG model, our implementation of negative sampling suppresses training sample "collisions." For a given training sample, a product pair can be either a positive or a negative sample, but not both at the same time.

Step II: Latent Product Attributes

The latent product attributes, derived for each product j, are based on the notion that similar products occur together in shopping baskets. For a given pair of a center product (ce) and a context product (co), the (scalar) similarity score between the two products is defined as the dot product of the products' latent attribute vectors, $score_{ce,\,co} = v_{ce} \times w_{co}$. v_{ce} and w_{co} are vectors of size $1 \times L$ and $L \times 1$. In prior NLP studies, common values for L range between 100 and 1,000, depending on the size of the vocabulary and the text corpus (Mikolov et al. 2013). In the original SG model, the total classification (log) loss for a given center product is defined as

$$Loss_{ce} = -\log\sigma(score_{ce,\,co}) - \sum_{k \in \{neg.\ samples\}} \log\sigma(-score_{ce,\,k}), \tag{1}$$

where $\sigma(x) = (1 + \exp\{-x\})^{-1}$ (Mikolov et al. 2013). The SG model is a binary classifier, so the loss function is often called binary cross entropy loss. The loss function consists of two additive components that are functions of the product vectors: the loss that the center product occurs together with the positive sample and the loss that the center product does not occur in a basket with a set of negative samples. Note that the SG model uses two representations for each product, v for center products

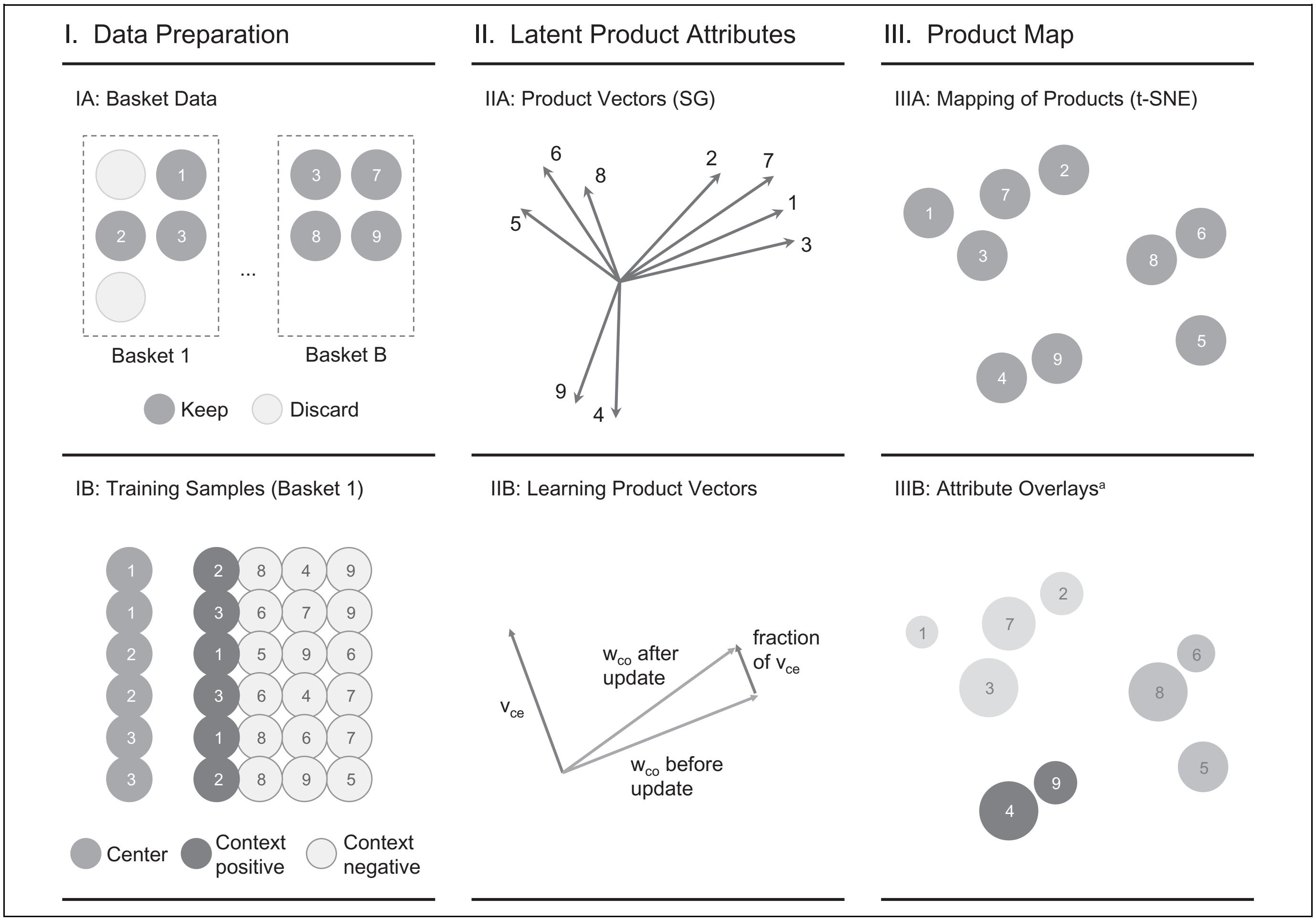

Figure 1. Overview of P2V-MAP.
[a]For example, brand as color, or market share as bubble size.

and w for context products. Mitra et al. (2016) consider how to interpret the different representations in NLP and find that vector products between words using representations from the same embedding ($v_{w1} \times v_{w2}$) are high for semantically similar words, but those between word vectors from different embeddings ($v_{w1} \times w_{w2}$) are high for words that co-occur often in the training data. The product map should primarily reflect product similarities, so we use v as input for t-SNE. The context embedding w is the key in analyzing complementarity. For complements j_1 and j_2, similarity is only a necessary condition and is not sufficient; similarity scores ($v_{j1} \times v_{j2}$) and co-occurrence scores ($v_{j1} \times w_{j2}$) must be large.

In addition to the modifications in the data preparation, we extend the SG score function

$$\text{score}_{ce,\ co} = v_{ce}\ w_{co} + \beta_0. \qquad (2)$$

The additional bias term β_0 captures the base probability that a product-center pair occurs in the data. It scales and centers product vectors (on average) around 0 (for details, see Web Appendix A1). In the analysis of market structures, it is an advantage that products and latent attributes are more comparable. Further extensions of the score function (i.e., product-specific bias terms, price as a time-specific covariate) did not change the substantive results (for details, see Web Appendix A2), so we use the more parsimonious score function in Equation 2.

The product vectors are derived by minimizing the loss defined across all training samples in the training data. The training data can contain billions of training samples, so the total training loss across the complete training data set comprises many individual loss terms, as defined by Equation 1. We iterate through all training samples and update the parameters by stochastic gradient descent (SGD) such that each parameter is changed by an amount proportional to the gradient of the loss for a set of randomly chosen samples (Hertz, Krogh, and Palmer 1991). We implement the modified SG model in TensorFlow (Abadi et al. 2015). Web Appendix A3 discusses the link between the SG model and neural networks. TensorFlow's reverse mode automatic differentiation for gradient operations avoids the need to derive the partial derivatives required for SGD. To explicate the intuition for deriving product vectors, we discuss the SGD update equation for the context product vectors,

$$w_{co}^{(new)} = w_{co}^{(old)} - \alpha\, \partial_{w_{co}} \text{Loss} = w_{co}^{(old)} - \alpha \{ \sigma(\text{score}_{ce,\ co}) - d_{pos} \} v_{ce}. \qquad (3)$$

α is the (nonnegative) learning rate and d_{pos} equals 1 if a context product is a positive sample and 0 otherwise. $\sigma(score_{ce,\ co}) < d_{pos}$ for positive samples. Therefore, the latent attributes for a center product in a given training sample are updated by adding a portion of the context product's latent attribute vector to the product vector of the center product (Figure 1), proportional to the classification error $\sigma(score_{ce,\ co}) - 1$. For a negative sample, updating achieves the opposite effect as $\sigma(score_{ce,\ co}) > d_{pos}$. The product vectors between positive samples and a center product are therefore made more similar to each other; the product vectors of negative samples and a center product are made more different. Updating the product vectors for all training samples and iterating through all samples n_{iter} times yields a solution in which the gradients diminish and the latent product vectors converge.

Using latent product attributes facilitates the automation of mapping market structures because it eliminates the need to collect and maintain master data or to manually label product attributes. That the SG model can be updated incrementally as new market basket becomes available simplifies scaling P2V-MAP. Latent properties also reduce the input data for the mapping step from a $J \times J$–dimensional distance or similarity matrix to a $J \times L$–dimensional attribute matrix; this data "compression" reduces time and space complexity in the mapping step and allows the mapping of a larger number of products. Another useful property of the SG model is that product embeddings produced by the SG model simplify nonlinearities in data. The linearity of product vectors means that linear vector arithmetic (e.g., addition, subtraction) works in the embedding space. Studies in NLP, for example, showed that the most similar word vector to $v_{king} + v_{woman} - v_{man}$ is the vector for queen v_{queen}. Vylomova et al. (2015) analyze the generalizability of word vector differences and conclude that many morphosyntactic and morphosemantic differences between words (e.g., synonymity, tense, count) are preserved in vector differences; we provide empirical evidence of embedding linearity using grocery data. Many dimensionality reduction techniques are based on Euclidean distances, so the output of the SG model should be very suitable in such cases. In addition, the linearity of product vectors makes it possible to map products for which no sales data are available (for two examples, see the "Mapping New Products" section).

Step III: Mapping of Products and Attribute Overlays

In the third step, we map products on the basis of their pairwise similarities, as captured by the center embedding v. The two-dimensional product map is the result of t-SNE dimensionality reduction applied to v or, formally, using the transformation $v \in R^{J \times L} \rightarrow y \in R^{J \times 2}$. A normalized Gaussian kernel models similarities between products in the latent attribute space, and a Student's t-distribution with one degree of freedom does so in the two-dimensional space. Noting the locations of products in the high-dimensional latent attribute space, we train the two-dimensional t-SNE representation y by minimizing the Kullback–Leibler divergence (i.e., the relative entropy) between the two probability distributions using stochastic gradient descent (Kullback and Leibler 1951; Van der Maaten 2014). To address the problem that t-SNE is not guaranteed to converge to a global optimum, we run t-SNE 100 times and pick the run with the lowest Kullback–Leibler divergence (Van der Maaten 2014). Before the t-SNE step, we L2-normalize product vectors, and for large embedding sizes ($L > 50$), we apply principal component analysis (PCA) ($n_{dim} = 30$) to the embedding matrix to reduce noise in the latent attributes.

t-SNE evaluates pairwise similarities between products, so the characteristics that determine product proximity can vary between clusters. This characteristic is required for mapping market structures across several product categories simply because a variety of product attributes can govern product relationships within different product clusters. While the two latent dimensions of the t-SNE map do not have a global meaning, map overlays allow us easily to discern what drives the similarity of products within a cluster. Product attributes added to map excerpts in the empirical application include price and brand. To further assist in its interpretation, additional information can be added to the market structure map, such as adjusting the size of bubbles (e.g., to symbolize market share). Product complementarity, captured by high co-occurrence scores $v \times w$, can be encoded in colored edges between bubbles (see the visualization of competition in Ringel and Skiera [2016]). Depending on the particular use case, additional data sources (e.g., loyalty card data) might further enrich the product map. As in the first two steps, the additions of different kinds of attribute overlays can be automated, for example, in an interactive dashboard.

t-SNE is a good choice for mapping products based on their latent product attributes because the heavy tails of the Student's t-distribution in the two-dimensional space allow dissimilar products to be farther apart in the reduced map, which is "desirable because it creates more space to accurately model the small pairwise distances (i.e., the local data structure) in the low-dimensional embedding" (Van der Maaten 2014, p. 3225). Compared with other algorithms, t-SNE produces more distinct clusters (Van der Maaten and Hinton 2008), which simplifies the interpretation of the product map. The tight, widely separated map clusters also make it easier for clusters to move around relative to each other. This allows t-SNE to derive a good global organization of product clusters while modeling within-cluster product similarity, according to small pairwise distances. Finally, the computational complexity of t-SNE, given n products, is $O[n \log(n)]$, so it scales better with growing product assortments than other dimensionality reduction techniques. The complexity of multidimensional scaling, for example, is $O(n^3)$. The ability to map millions of items (Van der Maaten 2014) allows for a wider applicability of our approach. To reduce run time and ensure that product maps are comparable across multiple runs, t-SNE can be initialized with the previous map results.

Simulation Study

Simulation Setup and Expected P2V-MAP Results

Verifying a new method for analyzing relationships between products in an empirical context is challenging because the true market structure is typically not known. In fact, studies are often limited to evaluating the face validity of empirical findings. In validating P2V-MAP, we leverage the fact that our approach requires only market basket data. Researchers have developed models to study category (e.g., Manchanda, Ansari, and Gupta 1999) and product choice (e.g., Fader and Hardie 1996), so the data-generating process (DGP) of market basket data is well understood. This makes it possible to use a simulation study to validate P2V-MAP, benchmark it against alternative methods for mapping market structure, and study the characteristics of the proposed method in a controlled environment (e.g., sensitivity to hyperparameters). Specifically, we generate data with a known market structure, apply P2V-MAP to these data, and compare the results with the simulated market structure. We withhold all information pertaining to market structure (e.g., true product categories, product relationships) and feed only market basket data to P2V-MAP. Comparing the P2V-MAP output with the parameters of the DGP yields a comprehensive test and validation of the proposed approach. The simulation study also provides the basis for the additional validation in the empirical application, where we show that P2V-MAP results converge with and extend the results of state-of-the-art methods for modeling category and product choice. We next describe the simulation setup and discuss how different components of the DGP relate to the output of P2V-MAP. A parsimonious simulation setup simplifies the discussion of the results while ensuring that the data reflect basic patterns pertaining to market structure.

In literature, a typical way to simulate consumer purchasing decisions in grocery retailing is a sequential process (Neslin and Van Heerde 2009). First, consumers select the categories within which they want to make a purchase. A category is defined as a set of products that are used interchangeably (e.g., coffee, laundry detergent). For each category, consumers then choose a product. We follow this approach in simulating market basket data. In week t, consumer i chooses product j from category c. Each category consists of $J^{(c)}$ products, and a product belongs to exactly one category. Not all categories have to be purchased in a given week. For modeling category purchase incidence, we use the multivariate probit model (MVP; Manchanda, Ansari, and Gupta 1999). This model allows for more than one category to be purchased simultaneously. Given a latent variable $z_{ict} = \gamma_c + \varepsilon_{ict}$ that depends on the base utility of category purchase γ_c and a (correlated) error term $\varepsilon_{ict} \sim MVN(0, \Omega)$, category purchase incidence is given by

$$y_{ict} = 1 \text{ if } z_{ict} > 0 \text{ and } y_{ict} = 0 \text{ otherwise.} \tag{4}$$

The correlation matrix Ω captures (random) purchase incidence correlation across categories. If the correlation is positive, two products are purchased together more often; negative correlations result in the opposite (Manchanda, Ansari, and Gupta 1999). To model (discrete) product choice within each category, we use the multinomial probit model (MNP; Chintagunta 1992a). The utility of product j for consumer i in week t is $u^{(c)}_{ijt} = \alpha^{(c)}_{ij} - \beta^{(c)} p^{(c)}_j + \varepsilon^{(c)}_{ijt}$. It is the sum of a base utility $\alpha^{(c)}_{ij}$, the (dis)utility incurred by paying price $p^{(c)}_j$ (multiplied by price sensitivity $\beta^{(c)}$) and a random error term $\varepsilon^{(c)}_{ijt} \sim N(0, \sigma^{(c)})$ that captures uncontrollable factors that drive product purchases. We assume that $\alpha^{(c)}_{ij} \sim MVN(0, \Sigma^{(c)})$, with a category-specific covariance matrix $\Sigma^{(c)}$. Consumers buy exactly one product per category and week (the one that offers the highest utility), that is,

$$y^{(c)}_{ijt} = 1 \text{ if } u^{(c)}_{ijt} > u^{(c)}_{ikt} \;\forall k \neq j \text{ and } y^{(c)}_{ijt} = 0 \text{ otherwise.} \tag{5}$$

From these models, we first draw the categories that consumers purchase in a given week and then sample the products consumers select in those categories. One caveat of this approach is that two products from the same category are never purchased together; therefore, the identification of categories would be trivial. In addition, market basket data show that consumers frequently buy more than one product per category on a given shopping trip, so we sample two product purchases in each category for 50% of all category purchases. We use $C = 20$ categories and $J^{(c)} = 15$ products per category. The MVP correlation matrix Ω has a block-diagonal form (four blocks with four categories and two blocks with two categories each), with manually specified positive and negative correlations of varying magnitude. This captures a variety of potential category relationships, including substitution, complementarity, and purchase coincidence. For each category, we compute $\Sigma^{(c)} = (\tau^{(c)} I^{(c)}) \Omega^{(c)} (\tau^{(c)} I^{(c)})$, where $\tau^{(c)}$ is the standard deviation of the product preference heterogeneity, $I^{(c)}$ is the $J^{(c)} \times J^{(c)}$ identity matrix, and $\Omega^{(c)}$ is a correlation matrix. To build $\Omega^{(c)}$, we sample (positive) partial correlations from Beta(.2, 1). Although most correlations are small (median = .11), 10% of the values are still larger than .54, so this setup enables us to simulate different degrees of product competition (i.e., similar preferences) within categories. Further details on the simulation study and parameter settings (including specific numerical examples for Ω and $\Omega^{(c)}$) are available in Web Appendix A4.

The final data set contains 1,877,991 shopping baskets that contain at least one product. We apply P2V-MAP to the simulated data and compare the DGP components with the P2V-MAP results; Table 1, Panel A, summarizes our expectations (E1–E7). Two products from the same category are more similar than two products from different categories (because they are substitutes), so we expect higher similarity scores ($v_{j1} \times v_{j2}$) within categories (E1). For the product map, this means that products from the same category should form clusters (E2). Within each cluster, products with higher covariances $\Sigma_{j1,\,j2}$ should be closer to each other than products with lower covariances because such products have similar utilities across consumers and compete more directly with each other (E3). Ω

Table 1. Simulation Study Expectations and Benchmarking Results.

A: Expected Relationships Between P2V-MAP Outputs and Simulation Parameters

Number	Model	Variables	Expectation
E1	MNP	Product similarity	Higher within true categories
E2	MNP	Product similarity	True categories form clusters
E3	MNP	Product similarity scores and product covariances (Σ)	Positive correlation
E4	MNP	Product map distances and product covariances (Σ)	Negative correlation
E5	MVP	Co-occurrence scores and category correlations (Ω)	Positive correlation
E6	MVP	Category similarity scores and category correlations (Ω)	Positive correlation
E7	MVP	Category map distances and category correlations (Ω)	Negative correlation

B: Method Comparison Results for Three Benchmarking Metrics

Method	Input Data for Mapping Step	Mapping Technique	Silhouette	Adjusted Mutual Information	Nearest Neighbor Hit Rate
P2V-MAP	Product vectors	t-SNE	**.350** (.004)	**.823** (.003)	**.696** (.003)
MDS-PV	Product vectors	MDS	.080 (.007)	.640 (.007)	.465 (.005)
PCA-PV	Product vectors	PCA	.049 (.002)	.517 (.004)	.388 (.003)
TSNE-CONF	Confidence	t-SNE	.056 (.006)	.692 (.005)	.486 (.003)
MDS-CONF	Confidence	MDS	−.038 (.029)	.486 (.028)	.368 (.026)
PCA-CONF	Confidence	PCA	−.203 (.000)	.149 (.001)	.153 (.000)
Random	—	—	−.196 (.003)	−.002 (.002)	.046 (.001)

Notes: Scores are averages over 30 bootstrap replications of the simulated data set with standard errors in parentheses; boldface indicates best approach per metric.

directly models category purchase coincidence, so Ω and the co-occurrence scores between average product vectors per category ($v_{c1} \times w_{c2}$) should have a positive correlation (E5). Categories that are bought together should also be similar to each other (because the same consumer buys them in the same shopping trip), so we expect a positive correlation between the similarity scores ($v_{c1} \times v_{c2}$) and Ω (E6). In line with our expectations about product proximity, similar categories should be closer to each other on the product map than dissimilar categories (E7). Given the positive correlation between category similarities and Ω, along with the positive correlation between similarities and cluster proximity, we expect a negative correlation between map distances and $\Sigma^{(c)}$ (and $\Omega^{(c)}$) (E4). In other words, products with similar utilities compete with each other and should be close to each other on the product map.

Simulation Results

To analyze similarities between products and categories, we first calculate the average similarity between all category pairs. For this, we average product vectors within each category and compute the dot product of the category vectors for all category pairs. Figure 2, Panel A, visualizes the results in the heatmap. In support of E1, within-cluster similarities (.42) are higher than similarities across categories (<.01). Categories with correlations of 0 show a similarity close to 0, and we observe positive (negative) similarities for categories with positive (negative) correlations. The correlation between similarity and Ω (for nonzero values in Ω) is positive ($r = .93$, $p < .01$). This confirms the expected positive relationship between cluster similarities and Ω (E6). The same is true for the relationship between product similarities and covariances $\Sigma_{j1,\ j2}$ (E3), for which we observe a significant positive correlation ($r = .76$, $p < .01$). The co-occurrence scores that are visualized in the heatmap in Figure 2, Panel B, look very similar to the similarity scores (Panel A). The most notable differences are the co-occurrence values within categories (diagonal), which are lower than within-category similarities. While products from the same category should have high similarities, products might co-occur more or less often with products from the same category. This illustrates the benefit of using two product embeddings, v and w (see Web Appendix A5 for a study of v and w in the context of the empirical application). The positive correlation between the co-occurrence scores and Ω ($r = .97$, $p < .01$) confirms E5.

Figure 2, Panel C, depicts the product map produced by P2V-MAP. Products from the same categories form clusters that are well separated (E2). High correlations in the MVP drive product similarity and can lead to merged product clusters, even though products belong to different categories (e.g., categories 17 and 18, categories 19 and 20). A real-life example of this could be pasta and tomato sauce. Next, we evaluate the correlations between within-cluster product distances and Σ and between cluster distances (cluster centers) and Ω. In support of E4 and E7, we find significant negative correlations for product distances ($r = -.46$, $p < .01$) and category distances ($r = -.79$, $p < .01$), respectively. The results confirm that similarity scores, co-occurrence scores, and the locations of products and clusters on the market structure map (Euclidean distances) capture simulated market structure.

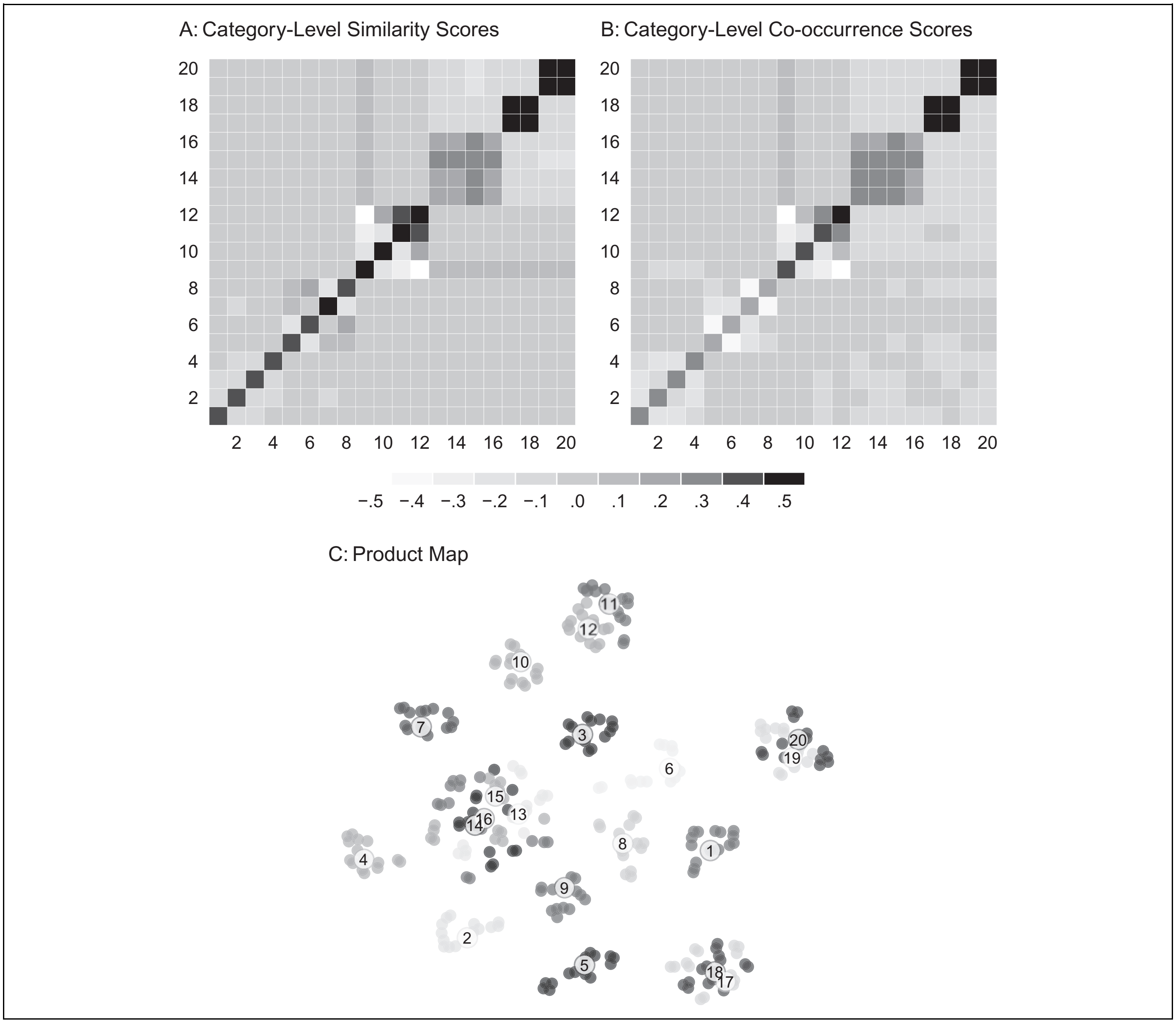

Figure 2. P2V-MAP uncovers simulated category structure.
Notes: Replication 1 of 30; shading indicate the true (simulated) product categories.

Method Comparison and Benchmarking

To benchmark P2V-MAP, we compare six approaches to mapping market structures. These approaches are combinations of three dimensionality reduction techniques and two types of input data derived from market basket data (Table 1, Panel B). The dimensionality reduction techniques are PCA, multidimensional scaling (MDS; Gower 1966), and t-SNE (Van der Maaten 2014). The implementations are based on Python's scikit-learn library (Pedregosa et al. 2011). These are applied to product vectors or confidence $p(j_1|j_2)$, which is the conditional probability that a consequent j_1 occurs in a basket if antecedent j_2 is in the basket (Blattberg, Kim, and Neslin 2008). For MDS, we calculate Euclidean distances between products (we set the diagonal of the confidence matrix to 0). A comparison with a product map based on a two-dimensional SG embedding is available in Web Appendix A6.

We evaluate the six methods using three metrics: The silhouette score (SIL; Rousseeuw 1987),

$$\mathrm{SIL} = \frac{1}{J} \sum_j \frac{b(j) - a(j)}{\max[b(j),\ a(j)]}, \tag{6}$$

evaluates the structuredness of the product map based on the average distance of product j to all other products in the same category, a(j), and the average distance of j to all products in the closest category on the product map, b(j). This metric is a good proxy variable for measuring how easily humans can interpret the map. Values range between −1 and 1. Negative values indicate that samples have been assigned to the wrong cluster, and values near 0 indicate overlapping clusters. The adjusted mutual information score (AMI; Vinh, Epps, and Bailey 2010),

$$\mathrm{AMI} = \frac{\mathrm{MI}(X,\ Y) - E[\mathrm{MI}(X,\ Y)]}{[H(X) + H(Y)]/2 - E[\mathrm{MI}(X,\ Y)]}, \tag{7}$$

is based on the mutual information score $\mathrm{MI} = \sum_{x \in X} \sum_{y \in Y} p(x,\ y) \log[p(x,\ y)/p(x)p(y)]$ and quantifies the agreement between the true category labels X and the predicted cluster labels Y; E[MI(X, Y)] is the expected mutual information, H(X) and H(Y) are the label entropies, p(x, y) is the joint probability distribution of X and Y and p(x) and p(y) are the marginal probability distribution of X and Y. If X and Y align perfectly, the AMI takes a value of 1 and a value of 0 is expected for a random cluster assignment. To predict cluster labels, we use k-means clustering on the product map coordinates (k is set to C). The AMI measures how well the product map recovers the true product categorization. The hit rate (HR),

$$\mathrm{HR} = \frac{1}{J} \sum_{j} f[j,\ c(j)], \tag{8}$$

calculates what fraction f[j, c(j)] of the $J^{(c)} - 1$ nearest neighbors belongs to the same category as the reference product, averaged across all products. The interpretation of the HR is similar to the mutual information score; however, by calculating the fraction of nearest neighbors, we avoid the additional clustering step.

For the benchmarking, we create 30 subsets of the original simulated data set through bootstrapping, each contains approximately 1 million baskets. Table 1, Panel B, contains the mean scores and standard errors of the three metrics across the 30 replications; Figure 3 depicts the resulting maps (replication 1) for all six combinations of input data (rows) and dimensionality reduction (columns). P2V-MAP is an unsupervised learning approach and the mapping between product co-occurrences and map coordinates is nonparametric, so we report benchmarking scores on the training data. P2V-MAP produces market structure maps with the best scores on all three evaluation metrics, followed by MDS-PV and TSNE-CONF. Note that running P2V-MAP on randomized data produces scores similar to the random baseline (for details, see Web Appendix A7). We find that t-SNE yields a two-dimensional representation of the product space that contains clearly disjointed clusters. Using t-SNE avoids an elliptical map shape, which is common for MDS representations (Ringel and Skiera 2016), and the formation of product lumps, as in the case of PCA. This simplifies the interpretation of the product map. Both components of P2V-MAP, product vectors and t-SNE, significantly improve the quality of the product map. On average, using product vectors instead of confidence increases SIL by .221, AMI by .218, and HR by .181. All three dimensionality reduction techniques rely on Euclidean distances between products, so the linearity of product vectors as discussed in the description of our mapping approach is likely the reason for improved product maps. The score improvements for using product vectors instead of confidence are smaller for t-SNE than for PCA. A possible explanation for this is that t-SNE is known to be capable of mapping nonlinear data to some extent because it considers only (Euclidean) distances between nearest neighbors (Van der Maaten and Hinton 2008). The improvements that result from replacing MDS with t-SNE are of a similar magnitude (SIL: .182, AMI: .195, HR: .175). The results suggest that PCA and MDS can be used to analyze market structure in that the scores for all approaches are higher than the scores for the random baseline, but P2V-MAP outperforms all other approaches because it generates the most structured product map and performs significantly better on all benchmarking metrics. Web Appendix A8 explicates how hyperparameter settings affect the performance of P2V-MAP. We comment on the stability of P2V-MAP results in Web Appendix A9.

Empirical Application

Implementation Details

We empirically evaluate the applicability of P2V-MAP with an analysis of market structures using basket data collected in a large German city over 12 months from 147 stores of the city's largest grocery retailers (Table 2, Panel A). The data consist of 3,386,500,350 training samples (derived from 73,048,605 shopping baskets) and J = 30,763 products (with $n_{min} > 100$) from 133 product categories. The product assortments are very similar across stores, as 87.0% of all products are listed in every store. The comparison of our data set with the English Wikipedia data set, as is often used in NLP, suggests that our approach scales well to larger data that contain hundreds of billions of training samples for millions of products (Table 2, Panel B).

For the SG model, we use our customized implementation in TensorFlow,[1] which learns product vectors through adaptive moment estimation ("Adam"), with default settings (Kingma and Ba 2014). One key benefit of Adam is the automatic, adaptive computation of learning rates for each parameter. In addition to exponentially decaying averages of past squared gradients (e.g., Adadelta and RMSprop), Adam uses an exponentially decaying average of gradients to improve learning stability and speed. Adam's sparse gradients update product vectors only if products occur in a gradient descent batch, which improves the scalability of P2V-MAP. The t-SNE implementation uses the scikit-learn library (Pedregosa et al. 2011).

One essential aspect of implementing machine learning algorithms is the choice of hyperparameters (Table 2, Panel C). The learning rate is tuned automatically through Adam. The two most important hyperparameters are the number of latent dimensions L and the perplexity in t-SNE (Web Appendix A7). A relatively large number of latent attributes, L = 300, makes it possible to encode many product attributes in the latent attribute

[1] See https://github.com/sbstn-gbl/p2v-map.

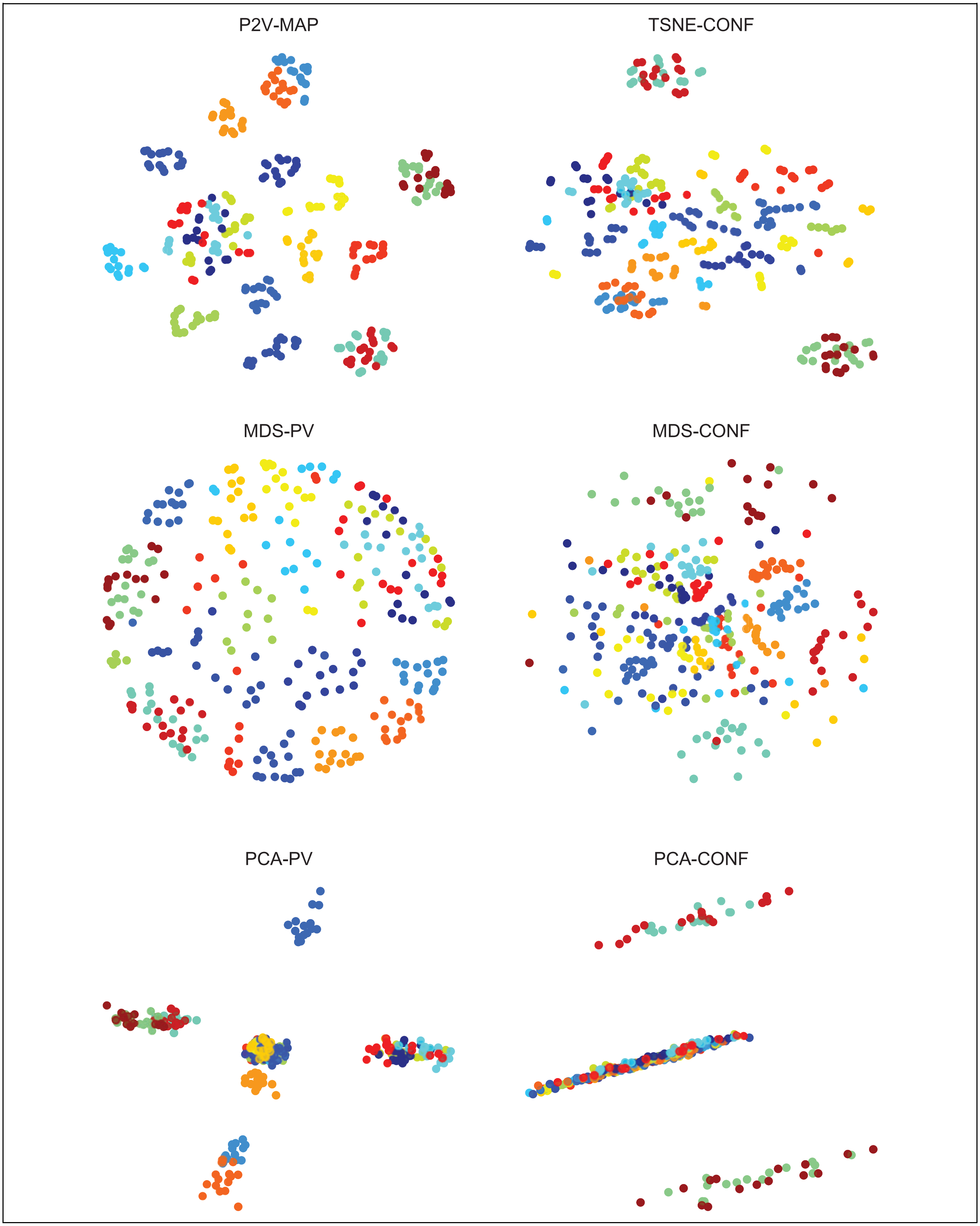

Figure 3. Comparison of product maps for six mapping approaches.

Notes: Replication 1 of 30; colors indicate the true (simulated) product categories.

Table 2. Basket Data Set Excerpt and Comparison of Data Set Characteristics and Hyperparameter Settings with NLP Application.

A: Market Basket Data (Map Between Basket Hash and Product ID) Is the Input Data for P2V-MAP

Basket Hash	Product IDs
1	26765200019, 51914800017, 71061700015, 75665700013
2	10294300012, 18476500014, 79630500011, 18337300012, 79033200013, 16267600011, 16994500011, 18312600014, 22570200015, 30805100016, 74460900017, 83298600013, 10294300012, 14666300013, 80665400016, 84982500015, 30805100016
…	…
73,048,604	15383500014, 17628800019, 18336700011, 51925300018, 51948500013, 52587600010, 76499500015, 78778600010, 81674100010, 18407900012, 68144300019, 14690600011, 51914800017
73,048,605	10294300012, 16649000019, 16677200016, 18404800018, 69035800014, 77336100016, 81962100029

B: Data Set Characteristics and Comparison to NLP Data

Variable	Market Basket Data	Wikipedia Data
# of baskets/sentences	73,048,605	115,844,936
# of training samples	3,386,500,350	57,262,363,914
Size of assortment/vocabulary	30,763	1,357,520
Avg. # of product/word occurrences	11,665	1,699
Avg. # of products/words per basket/sentence	4.91	18.86
# of stores	147	N.A.
# of months	12	N.A.
# of categories	943	N.A.

C: P2V-MAP Hyperparameter Settings and Comparison to Typical Hyperparameter Settings in NLP Applications

Hyperparameter	Market Basket Data	NLP
# of latent dimensions L	300	100 –1,000[a]
# of negative samples n_{neg}	20	2–20[a]
Iterations n_{iter}	10	1–3[a]
Power pow	.75	.75[a]
Perplexity	30	5–50[b]

[a]Mikolov et al. (2013) is the source for the NLP hyperparameters.
[b]Van der Maaten and Hinton (2008) is the source for the NLP hyperparameters.
Notes: Wikipedia data statistics based on English Wikipedia dump (December 1, 2018); in preprocessing we remove punctuation, transform all words to lowercase, and base the vocabulary on words with at least ten occurrences. N.A. = not applicable.

space. Using more than 300 latent attributes does not lead to more (or different) product clusters. Regarding perplexity, a value of 30 (or larger) produced good results. For the number of negative samples, Mikolov et al. (2013) recommend values between 2 and 5 for large data sets and 5 and 20 for small data sets. Our data are not as vast as the English Wikipedia corpus, so we set the number of negative samples to $n_{neg} = 20$. We do not observe any overfitting in our implementation of the SG model, so for n_{iter}, large values are reasonable (compare the holdout loss curves in Web Appendix A10). In practice, we recommend monitoring changes in the (holdout) loss and product vectors and stopping training when loss and embeddings stabilize. For the data studied here, this is the case after iterating through all 3.4 billion training samples $n_{iter} = 10$ times.

Next, we discuss our findings regarding the category-level and product-level market structures. These illustrate the usefulness of P2V-MAP in a marketing context (Web Appendix A11 contains an additional method comparison, using the same six methods we compare in the simulation study). For each of the following subsections, we conduct an additional validation step. That is, we compare the results of P2V-MAP with the results of state-of-the-art econometric analyses, which we apply to loyalty card data from the same retailer. Loyalty card data are well suited for the validation tasks because users' loyalty card IDs allow us both to limit our analysis to category users and to account for (unobserved) heterogeneous product preferences. In line with the DGP of our simulation study, we use multivariate and multinomial discrete choice models to analyze consumer decisions. Note that the application of complex econometric models is feasible only if we restrict the analyses ex ante to rather small sets of categories or products (informed by the product map). In contrast, P2V-MAP produces very similar results without requiring additional constraints or assumptions.

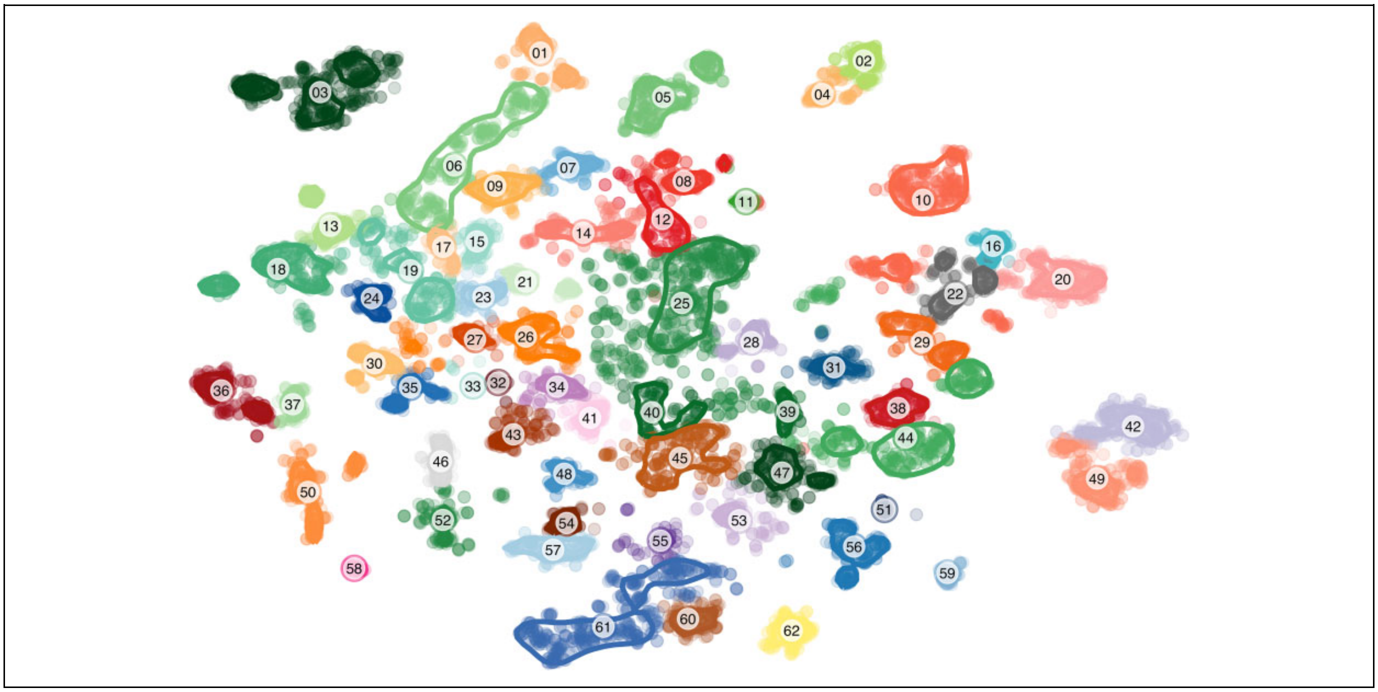

Figure 4. P2V-MAP product map for grocery retailer data.
Notes: The contours mark cluster regions that contain 90% of all products in the respective cluster; Web Appendix A12 describes the creation of the overlay and contains the cluster names.

Cluster-Level Market Structure

The product map we derive by applying t-SNE to the latent product attributes is depicted in Figure 4. We add an overlay that shows market clusters derived through k-means clustering (bubble colors and the corresponding numbers); these clusters are more meaningful than retailer categories (for details, see Web Appendix A12). Cluster contours indicate areas that contain 90% of a cluster's products. The overlay highlights one way to use the product map and enables us to check the face validity of the results. The two dimensions cannot have a single, global meaning across clusters, and both dimensions are normalized; therefore, we omit axis labels and values.

The most important feature of the product map is product proximity. Small Euclidean distances reflect high product similarity. First, we see that related products form clusters and related clusters build cluster groups. On the lower-right side of the map, we find nonfood products, which are grouped into the clusters body care (49), home care (42), and, in close proximity, baby products (59). On the top-right, beverages are located in a larger group of clusters, including soft drinks (38), juices (29), and waters (44), along with alcoholic beverages, such as beer (10), wine (20), sparkling wine (16), spirits and mixers (22). Tobacco products (31) are located next to alcoholic beverages. Sweets can be found at the bottom of the map with candy (60), cookies (57), and chocolate (61). Pet food (02: cat food, 04: dog food) forms a cluster group at the top right of the map. In the center of the map, we find a large group of breakfast products (25), with juices and smoothies (39), yogurt (45), and coffee (28) in close proximity.

Other product clusters are less obvious but still plausible. Products from different product categories form clusters if the products are complements, that is, they have the same use case (high similarity $v_{j1} \times v_{j2}$) and are frequently purchased together (high co-occurrence score $v_{j1} \times w_{j2}$). For example, on the left side of the map, products are grouped into different food and recipe clusters (e.g., 07: barbecue, 17: German, 18: Italian, 36: baking). Other clusters that are not category-specific are driven by distinct product attributes and include products designed to appeal to specific consumer segments (e.g., 26: organic, 30: vegetarian, 32: lactose free, 33: gluten free). Cluster 53 contains products for children (e.g., small juice boxes, snacks for school, bologna in the shape of teddy bears, Disney yogurt).

Larger clusters, such as yogurt and wine, show interesting substructures indicating different levels of competition that are driven by relevant product attributes and consumer preferences. Attribute overlays easily uncover such drivers. The map excerpts in Figure 5 contain yogurt and dessert products (Panel A), along with wine products (Panel B). Yogurt and dessert products are located in subclusters next to each other. In both subclusters, the brand attribute clearly delineates products. This is in line with prior research that points to the brand as one of the primary yogurt attributes for which consumers have very heterogeneous preferences (Chintagunta 1992b). In the wine cluster, regular wine, sparkling wine, and Piccolo bottles (200 ml or 250 ml) form distinct product groups. In each subcluster, product price is the main attribute that determines product location: lower-priced wine bottles (<€2 per .75 L bottle)

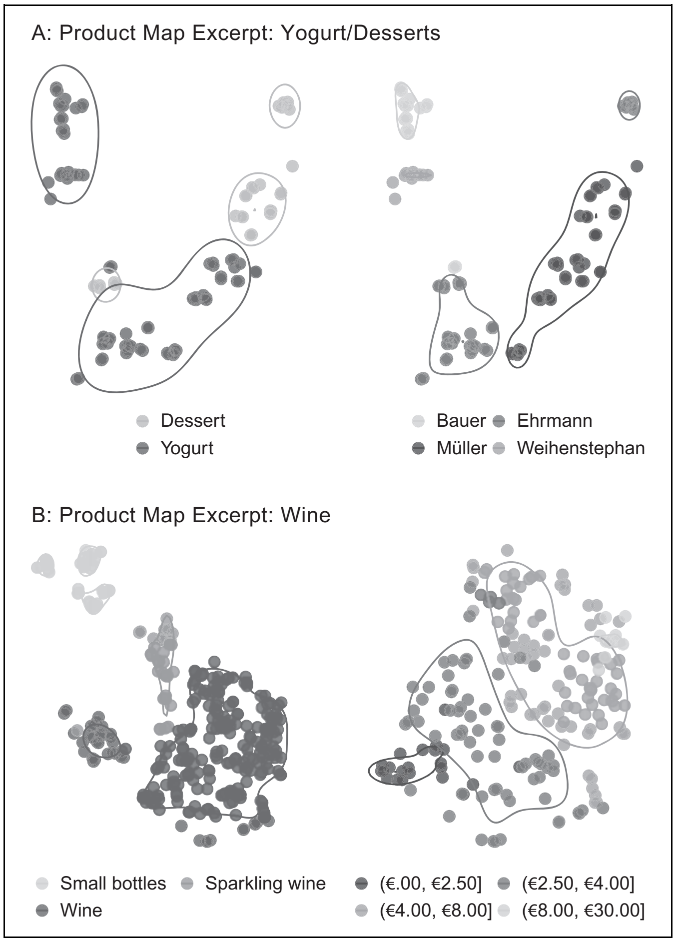

Figure 5. Map overlays reveal cluster-level drivers of market structure.

Notes: The contours mark cluster regions that contain 90% of all products in the respective subcluster, brand, or price range.

are located at the bottom left, and the price increases toward the upper right of the map except in multiple groups of typical price ranges for wine (e.g., €2–€4, €4–€10, and >€10). This result is supported by research that points to relatively high consumer loyalty to specific price tiers in the wine category (Jarvis and Goodman 2005). The two examples illustrate that the product map, in combination with simple descriptive overlays, allows us to infer drivers of market structures.

To further validate the results pertaining to cluster-level market structure, we take a closer look at the part of the product map that contains home care, body care, and baby products (Figure 6, Panel A). We employ loyalty card data and analyze (category) purchase incidence in six retailer categories c that are part of the three P2V-MAP product clusters: detergent, dishwashing liquid, deodorant, oral care, diapers, and baby food. For each shopping trip t, we know whether consumer i purchased a category ($y_{ict} = 1$) or not ($y_{ict} = 0$). To understand the relationship between the categories, we compute the tetrachoric correlations (Olsson 1979) of the binary category incidences in an MVP model (Manchanda, Ansari, and Gupta 1999). Keeping in mind the link between category-level purchase incidence and purchase correlation discussed in the simulation study, we expect positive correlations between all product clusters but higher correlations between categories within the same product cluster (e.g., detergent and dishwashing liquid in home care, oral care and deodorant in body care). Correlations in purchase incidence between categories from different product clusters (e.g., detergent in home care and deodorant in body care) should be lower. We include consumers with at least 3 but not more than 100 purchases in any of the six categories, which seems reasonable given the typical interpurchase times of these nonfood categories. The final sample consists of 251,125 shopping baskets of 3,823 loyalty card users with (on average) 76 shopping trips in 2016 (i.e., one trip every four to five days). In 20,895 baskets, at least one of the six categories was purchased. The 230,230 empty baskets ensure that the resulting correlations have a meaningful interpretation.

The heatmap in Figure 6, Panel B, visualizes the MVP correlations (confidence intervals [CIs] are derived through bootstrapping with 1,000 replications). We find values between 0 and .3. The correlations of categories within the P2V-MAP product clusters are higher than the correlations between clusters. For example, the correlation between detergent and dishwashing liquid is .275 with a 95% CI of [.254, .296]; the values between deodorant and oral care and between diapers and baby food are of similar magnitude and precision. In contrast, the correlation between detergent and deodorant is only .120 with a 95% CI of [.080, .161]. Thus, the value is not only significantly different from zero but also significantly lower than the within-cluster correlations. These results provide empirical support for the market structure derived by P2V-MAP using a different data source. They are also in line with the results of the simulation study that establishes a negative correlation between cluster map distances and purchase co-occurrence.

One interesting detail is the relatively high MVP correlation between the categories oral care and diapers (.128). That the correlation is significant and of similar magnitude to the correlations between home and body care categories might be surprising at first. In contrast, the product map positions baby products further from the home care and body care clusters. A closer examination of the data and the products in the baby cluster (Figure 6, Panel A, lower-left corner) reveals that the retailer's category definition leads to this result. The category oral care includes not only regular products for adults but also products specifically designed for babies and children (the six products slightly separated from the baby cluster). After splitting the category oral care into two subcategories (i.e., regular and kids) and repeating the MVP analysis, we find that oral care (kids) is strongly correlated with diapers and baby food, with significant correlations of .246 and .318, respectively (Figure 6, Panel C). Oral care (regular) and baby food, however, have a very low correlation of .006 (not significant). In this case, the retailer's category definition does not reflect the consumers' view of market structure; rather, it is merely driven by logistics and category management. Using P2V-MAP product clusters in the MVP yields more intuitive results because

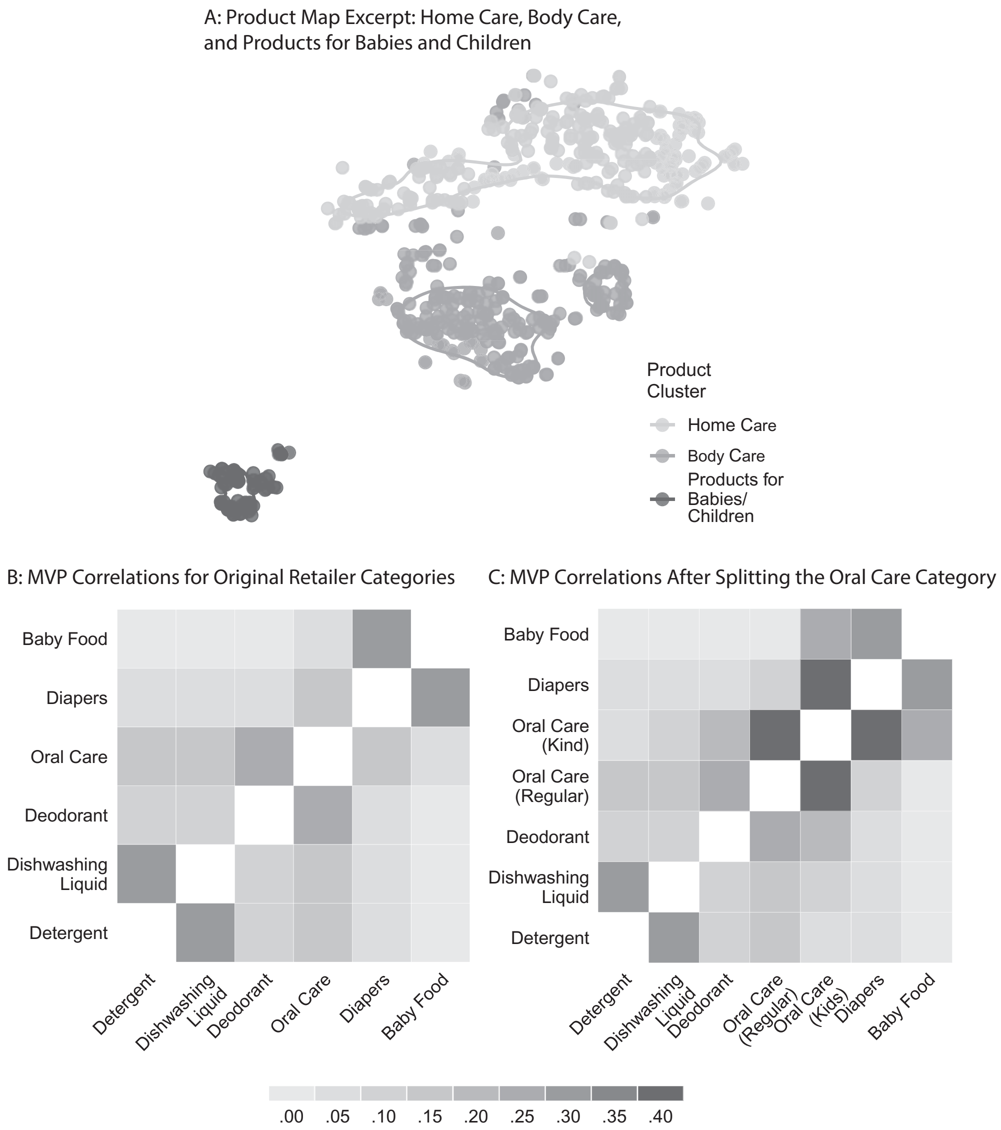

Figure 6. Results for category-level market structure.

Notes: The contours mark cluster regions that contain 90% of all products in the cluster.

P2V-MAP is based on cross-category information (extracted from market basket data) and no a priori assumptions on product relationship were made in deriving product clusters.

Product-Level Complementarity

Next, we turn to market structures at the product level. Several clusters contain complementary products; here, we present details for the barbecue cluster. One particular part of the barbecue cluster contains ingredients for "American" barbecue dishes such as hamburgers and hot dogs (Figure 7, Panel A). Hot dogs and hot dog buns are located on the left side, burger buns and beef patties on the right side. In between, we find condiments that can be used in hot dogs and hamburgers (e.g., sliced pickles, fried onions, sauces). The locations of the products on this map excerpt are intuitive and clearly point to the synonymity of products. For product complementarity, similarity is only a necessary condition and is not sufficient: (1) complements are located close to each other on the product map (high similarity), and (2) co-occurrence scores between complementary products are high. Figure 7, Panel B, therefore shows the co-occurrence scores for five products in the map excerpt. We find that scores are high between hot dog buns and hot dogs and between burger buns and beef patties. Both product pairs have high co-occurrence scores with sliced pickles but not with each other. This points to the complementarity of hot dog ingredients and burger ingredients.

As in the case of category-level market structure, we use loyalty card data in combination with an MVP to validate the output of P2V-MAP. Instead of analyzing the co-occurrence of

Figure 7. Complements and substitutes in the product cluster "American Food."

categories, we focus on joint purchases of products. Figure 7, Panel C, shows the tetrachoric correlations from an MVP analysis of the loyalty card data. The results are based on a random sample of 1,000 loyalty card users with 132,976 baskets; 3,319 baskets contain at least one of the products studied in this analysis. Purchases for hot dog and hamburger products are strongly correlated (hot dog: .651, 95% CI = [.593, .709], hamburger: .675, 95% CI = [.625, .725]), and sliced pickles "link" the two product groups through significant positive correlations (e.g., beef patties: .275, hot dog buns: .340). The correlations between hamburger and hot dog products are not significantly different from zero. When comparing the MVP results with the co-occurrence scores, we see that the two heat maps look remarkably similar. The positive correlation of .804 ($p < .01$) between tetrachoric correlations and P2V-MAP co-occurrence scores further supports the validity of the product map. Note that this also illustrates how P2V-MAP can complement more specific econometric methods in that it might guide data preparation and pruning steps before further in-depth analyses. A second test that analyzes to what extent the SG model captures the co-occurrence of products in shopping baskets can be found in Web Appendix A13; in this test, we remove items from shopping baskets and show that the SG model co-occurrence score predicts the missing items more accurately than metrics derived from market basket analysis.

Product-Level Competition

To illustrate how the product map facilitates the analysis of product competition, we first zoom in on the bottom-left corner of the sweets cluster (61). All 14 products in the map excerpt in Figure 8, Panel A, are 100-gram chocolate bars with a cocoa content above 50% (i.e., "dark" chocolate) with one exception: Lindt 99% 50 gram. We have made no a priori assumptions about these products being substitutes; instead of mapping the market structure according to a narrow set of products, we use

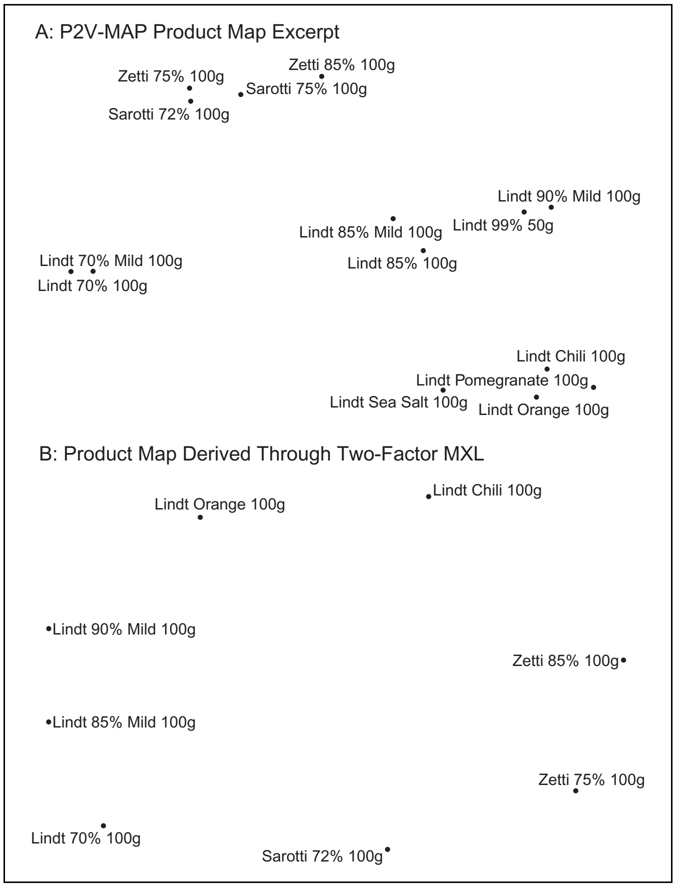

Figure 8. Product competition in the P2V-Map cluster "Dark Chocolate."

the shopping basket co-occurrence of all products in the retailer's assortment to derive the products' positions on the market structure map and then assessed their competitive relationships by interpreting the map ex post. The map excerpt does not contain any chocolate bars with a low cocoa content (e.g., white or milk chocolate), which shows that consumers do not view such chocolates as substitutes for chocolate bars that contain a higher percentage of cocoa. The chocolate subcluster has a clear structure; at the top, we find the brands Zetti and Sarotti, whereas the brand Lindt is predominantly located in a horizontal stripe in the middle. This separation offers face validity in that Zetti and Sarotti products are sold at a price of €1.29 and produced in East Germany, while Lindt chocolate is a Swiss product priced at €1.99. Across brands, we find that the cocoa content increases from left to right. An exception is the group of chocolates from Lindt at the lower right of the map excerpt; these chocolates have additional ingredients (e.g., orange, chili, sea salt). These products feature a cocoa content of above 50%, so their proximity to the other Lindt chocolate bars is reasonable. It is particularly noteworthy that the map contains several product pairs that are very close to each other—for example, Lindt 70% and Lindt 70% mild along with Lindt 85% and Lindt 85% mild. Consumers perceive these products to be very similar, as indicated by their proximity on the product map.

To validate the findings, we estimate a mixed logit (MXL) model (Train 2009) based on loyalty card data. This validation approach is in line with the simulation study that linked a higher map proximity to the larger covariance between product preferences in the MNP model. To make the model estimation feasible (e.g., size of the choice sets and the covariance matrix), we pick nine products that aptly represent the different types of products in the map excerpt: Lindt 70%, Lindt 70% mild, Lindt 85% mild, Lindt 90% mild, Sarotti 72%, Zetti 75%, Zetti 85%, Lindt Orange, and Lindt Chili. We draw a random sample of 500 consumers who made at least 4 but no more than 50 choices in the dark chocolate category in 2016. The final data set contains 4,215 purchases. Our goal is to understand the basic substitution patterns between products, so we estimate a model with a full covariance matrix for unobserved consumer heterogeneity (but without covariates). We infer the MXL correlations using maximum simulated likelihood (500 Halton draws) and compare them with the Euclidean distances between products on the product map. Preference correlations for close substitutes are high, so these products should be located next to each other on the map. The map distances and the MXL correlations show a significant negative correlation of $r = -.53$ ($p < .01$). A descriptive analysis of loyalty card data further supports these findings. We compare the Euclidean distances between products on the map with their Jaccard similarities. The Jaccard similarity is a popular measure in collaborative filtering (Blattberg, Kim, and Neslin 2008). A negative correlation of $r = -.79$ suggests that products with higher similarity are located closer to each other on the map (for details, see Web Appendix A14). Last, we train a model with a two-dimensional factor structure (Elrod and Keane 1995) and use the values of the factorized covariance matrix to plot a second product map (Figure 8, Panel B). The map clearly resembles the map derived through P2V-MAP in that brand and cocoa content are the most relevant factors. Products with additional ingredients are located somewhat separately.

Mapping New Products

Most techniques for visualizing market structure depend on past data (e.g., market basket data, consideration sets). It can be desirable to understand how products relate to other products in the market before data are available—for example, before a new product is introduced. Thus far, we have discussed the case of using P2V-MAP to map products for which purchase data are available. The linearity of product vectors makes it possible to map products that have not been observed in the past. Product vectors for new products are imputed through a linear combination of existing products. Next, the product is mapped just as if a product vector were available in the first place. Equation 9 depicts an example of how product vectors are imputed. Given the product vectors of Mars chocolate bars, Snickers chocolate bars, and Snickers ice cream bars, the vector for Mars ice cream bars is derived by

$$v_{\text{Mars}}^{\text{Ice cream bar}} = v_{\text{Mars}}^{\text{Chocolate bar}} + \left(v_{\text{Snickers}}^{\text{Ice cream bar}} - v_{\text{Snickers}}^{\text{Chocolate bar}}\right). \tag{9}$$

The difference between the product vectors for Snickers ice cream and Snickers chocolate captures the attribute "frozen." Adding this characteristic to the vector for Mars chocolate results in a "frozen" version of the chocolate bar (i.e., Mars ice cream). Given the availability of three product vectors A, B, and C and the fact that the difference between B and C captures an attribute that could be added to A, the product vector for D can be derived through the linear combination of A, B, and C. The vector for D can then be used in mapping market structure. Web Appendix A15 contains 16 examples to which this approach applies.

To verify that imputed product vectors can be successfully used for mapping market structure, we present two examples: Mars ice cream (see Equation 9) and Landliebe strawberry yogurt (= Landliebe cherry yogurt + Schwartau strawberry jam − Schwartau cherry jam). We first train two product embeddings: PE1 is based on all available market basket data (the regular P2V-MAP approach); for PE2, we remove all baskets that contain products D from the data set (Mars ice cream and Landliebe strawberry yogurt) and infer the product vectors for D, as illustrated in Equation 9. We produce the product maps by running t-SNE on both embeddings and compare the nearest neighbors of D. These should be similar in the two product maps. Distances between products on the final t-SNE maps are invariant under reflection and rotation. To ensure a similar orientation of the two maps and to simplify the comparison, we initialize the positions of the products on product map 2 with the positions from map 1, with the exception of product D, for which the locations are initialized with random starting values.

Figure 9 depicts the resulting product map excerpts around D. Map 1 is based on the original product vector for D (left), while map 2 is based on the imputed product vector (right). All four maps contain the 19 nearest neighbors to D. We label the products with integers according to their proximity to D on the map that is based on PE1 (1 being the closest to D). The color/shading labels indicate whether a product is among the 19 nearest products on map 1 (dark gray) or not (light gray). We see that the product maps based on PE1 and PE2 are very similar. With few exceptions, the 19 nearest neighbors from map 1 are also among the 19 nearest neighbors on map 2. The distances between corresponding products are very similar on both maps, and products form similar subclusters. Most interesting is that this is also true for product D. Regardless of which type of product vector is used for mapping (the SG model vector or the imputed product vector), product D occurs in the same map subcluster, and the nearest neighbors are very similar. This shows that products for which no data have been observed can be mapped successfully if it is possible to impute product vectors through a (linear) combination of other products. This sets P2V-MAP apart from other methods for market structure analysis.

Product Assortment Analysis Across Stores

Thus far, we have focused on using P2V-MAP in the descriptive analysis of market structures. In this section, we provide initial ideas for how P2V-MAP can be used to evaluate and possibly inform assortment strategies. For this, we first calculate the store-specific importance of each P2V-MAP product cluster. The product clusters reveal consumers' view of the market structure so the cluster-specific importance can be used to adjust product presentation or marketing activities. Another way to cater to consumer needs is to tailor the assortment depth to match the store-specific preferences of consumers. That is, we evaluate the product depth for each cluster-store combination and compare it to the store-specific cluster importance. This enables the retailer to strategically increase assortment depth in important clusters while reducing it in less important product clusters. The P2V-MAP product clusters yield a clearer understanding of consumer shopping behavior than retailer product categories because products are grouped into more meaningful clusters: some product clusters contain multiple product categories, whereas other product categories are spread over multiple clusters.

Given value sales per product cluster k and store s, vs_{ks}, we calculate for each store which share of value sales falls into a product cluster, vsh_{ks}. Then, we derive the relative deviation from the cluster-specific average across S stores, Δ_{ks}, such that

$$\Delta_{ks} = \frac{(vsh_{ks} - \overline{vsh}_k)}{\overline{vsh}_k} \text{ with } vsh_{ks} = \frac{vs_{ks}}{\sum_k vs_{ks}} \text{ and } \overline{vsh}_k = \frac{1}{S\sum_s \overline{vsh}_{ks}}. \tag{10}$$

Considering that vsh_{ks} can be interpreted as the importance of cluster k in store s, Δ_{ks} is a measure of the importance of clusters relative to the average cluster importance across stores. In Table 3, Panel A, we compare the Δ_{ks} values for 16 product clusters across three sample stores. We chose the stores because they differ in their location and clientele. Store A is located in a residential area with larger households, higher income, higher education, and low unemployment rates. Store B is also in a residential neighborhood, but the income and education rates are lower and unemployment rates are higher. Store C is in the city center, located among tourist attractions and office high rises. Although the stores' product assortments are similar (96.5% of the products listed in the store with the smaller product range are also listed in the store with the larger assortment), we observe interesting differences between the stores' sales across product clusters.

Consumers in Store A buy proportionally more organic products, vegetables, and salads, but fewer recipe mixes and meat products, such as barbecue meats and cold cuts. They prefer fresh juices/smoothies and wine over beer. More families with children live in the area around Store A, so its higher sales of baby products are reasonable. In nonfood clusters, we observe higher sales for home care products, as expected in a

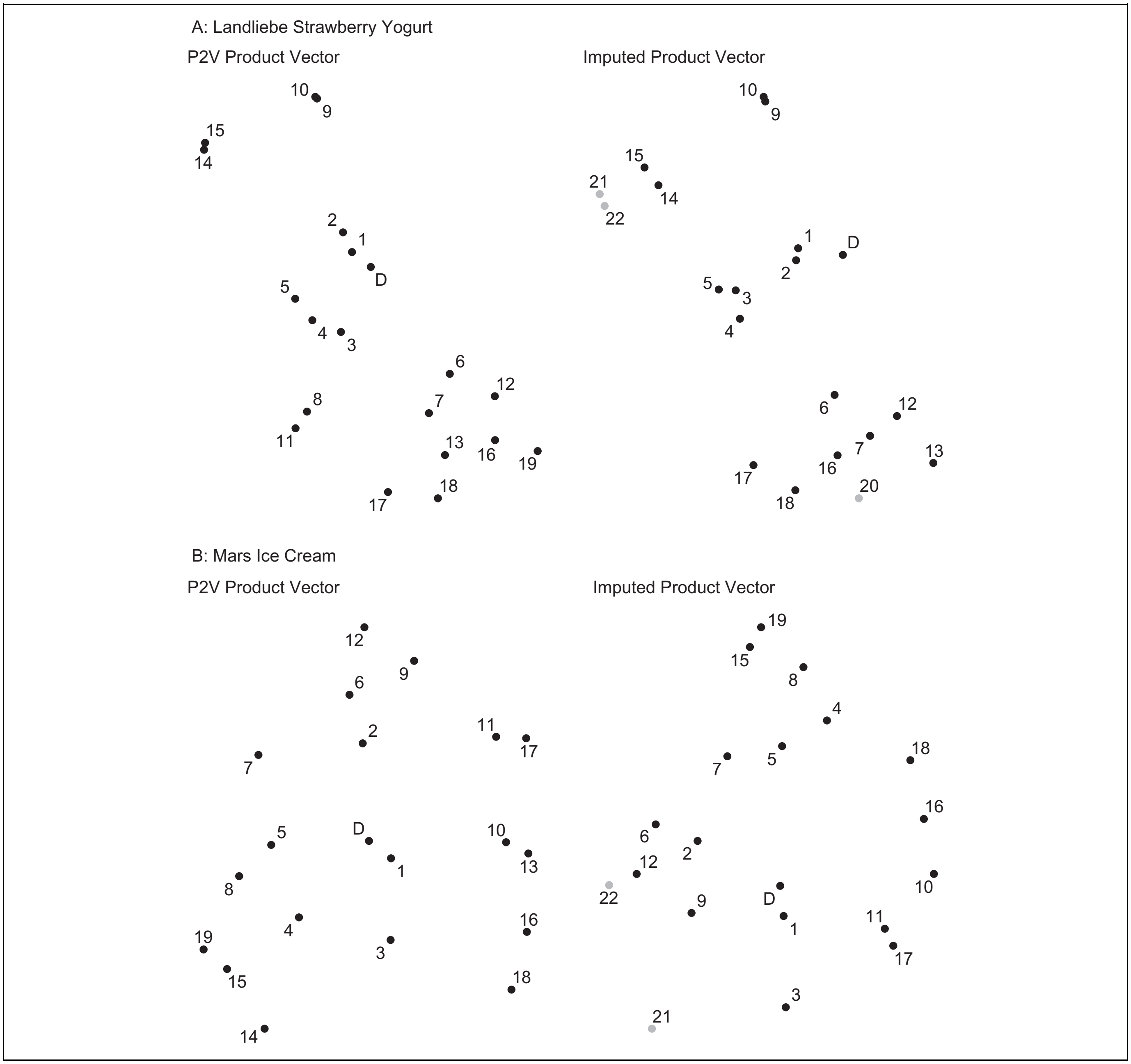

Figure 9. Results for mapping new products.

residential area. In contrast, in Store B, consumers' diets feature more convenience products, such as recipe mixes. Instead of international food (e.g., Asian food), these consumers prefer German food and meat products but not organic products, vegetables, or fruits. Store B's consumers buy less juice, and they prefer beer, coinciding with more sales of tobacco. For a store in a residential area, higher sales in the clusters home care and dog food are also plausible. Store C differs from both other stores in that snack products (lunch snacks and ice cream single packs) play a more important role because many shoppers are tourists or employees in nearby offices. Purchases of fresh juices and smoothies, wine, beer, and alcopops are higher in Store C. We find lower sales for baby products and nonfood items, such as home care and dog food, as expected for a store in the (nonresidential) city center. These findings have face validity and are in line with prior research on consumer shopping behavior. For example, higher income and higher education tend to encourage a diet that features more organic food and produce (Padel and Foster 2005), whereas lower income tends to imply lower sales of fresh fruits and vegetables and higher sales of meat and sugar-laden products (Anderson and Morris 2000).

To ascertain whether the retailer's assortment reflects this purchase heterogeneity, we analyze assortment depth, defined as the

Table 3. Store Assortment Case Study Results.

A: Deviation Value Share from Store Average–16 Clusters with Large Differences Between Stores

Product Cluster	Store A	Store B	Store C
Asian	50.8%	−17.5%	−26.3%
Baby Products	129.7%	49.6%	−44.6%
Barbecue	−23.8%	71.2%	−42.4%
Beer	−34.5%	31.2%	26.0%
Deli Counter (Meat & Cold Cuts)	−31.4%	52.8%	−44.6%
Dog Food	−41.7%	80.7%	−70.8%
Fresh Juices & Smoothies	58.0%	−47.1%	184.2%
German Food	−45.4%	31.0%	−53.7%
Home Care	40.6%	16.9%	−62.3%
Ice Cream (Single Packs)	−40.3%	36.8%	69.3%
Lunch Snacks	−38.0%	−28.2%	200.1%
Organic Products	149.9%	−53.5%	−2.6%
Recipe Mixes	−65.5%	38.6%	−65.1%
Tobacco	43.7%	79.0%	8.2%
Vegetables, Salads & Herbs	35.4%	−35.7%	−14.1%
Wine	15.0%	−21.4%	27.4%

B: Seven Product Clusters in Store C with the Highest/Smallest Differences Between Deviation Value Share and Deviation Assortment Depth

Product Cluster	Deviation Value Share from Store Average	Deviation Assortment Depth from Store Average	Fraction of Products Not Stocked in Store C	# Retailer Categories in Product Cluster	Avg. Cluster Importance Across Retailer Categories in Product Cluster
Lunch Snacks	+200.1%	+18.7%	3.8%	6	+40.8%
Fresh Juice & Smoothies	+184.2%	+30.1%	7.1%	2	+16.2%
Meat Snacks	+157.8%	+54.2%	.0%	1 (12.5%)	+6.3%
Ice Cream (Single Packs)	+69.3%	−26.6%	45.7%	1 (18.5%)	−1.2%
Protein and Cereal Bars	+105.5%	+31.2%	.0%	1 (14.7%)	+33.3%
Convenience Food	+56.9%	+9.7%	18.6%	18	+46.7%
Water	+42.2%	+7.0%	17.9%	1 (∼ 100.0%)	+38.0%
…					
Cake	−48.5%	−10.1%	35.7%	2	−31.4%
Soups	−38.4%	+2.8%	17.0%	2	−12.6%
Baby Products	−44.6%	−1.7%	21.4%	7	−32.5%
Barbecue	−42.4%	+1.3%	19.6%	20	+12.5%
German Food	−53.7%	−8.3%	26.4%	23	−4.8%
Easter Products	−40.4%	+5.9%	19.4%	2	−18.5%
Recipe Mixes	−65.1%	−10.0%	33.1%	4	−22.2%

Notes: The percentage values in parentheses for the column "# Retailer Categories in Product Cluster" indicate the fraction of the categories' products that belong to the product cluster (if value is 1).

number of distinct products in a given product cluster. First, we derive the fraction of a store's product assortment that belongs to cluster k. Just as we did for the value shares in Equation 10, we then calculate the relative deviation of this variable from the average across stores, Δ^{ad}_{ks}. Table 3, Panel B, suggests that the retailer's assortment depth does not reflect consumers' purchase behavior in many clusters, using the case of Store C (city center). The seven clusters with the highest (top) and lowest (bottom) differences $\Delta_{ks} - \Delta^{ad}_{ks}$ are depicted, along with the fraction of all available products in each cluster not stocked in Store C. We first note the higher sales and deeper assortments for the clusters of lunch snacks, fresh juices and smoothies, meat snacks, and protein and cereal bars. Store C stocks nearly all available products, so an additional increase in assortment depth would be difficult. In other clusters, however, higher sales stem from a narrower assortment. For ice cream single packs, convenience food, and water, there is the potential to list more products and better cater to consumer preferences by offering a wider product range. If adjusting the assortment depth is not possible, these insights could inform the choice of product presentation. The most important product clusters could be featured (more) prominently in the store and in secondary placements. In other clusters, Store C should consider a (further) reduction in assortment depth, such as soups, baby products, and recipe mixes.

Table 3, Panel B, underlines the importance of using P2V-MAP product clusters in place of the retailer's category-centric view of the market structure. To compare the results between the two ways of structuring the product assortment (P2V-MAP clusters vs. retailer categories), we count the number of distinct retailer categories c to which the products from a specific product cluster k belong. The only product cluster that exactly aligns with a retailer product category is water; in all other cases, product clusters contain several categories (e.g., German food: 23 categories) or a small subset of the products in a category (e.g., meat snacks: 12.5% of all category products). $E_c[\Delta_{ks}]$ is the weighted average of the cluster importance Δ_{ks} across the retailer categories present in cluster k, using category value sales as weights. For several product clusters, the cluster importance Δ_{ks} differs substantially from $E_c[\Delta_{ks}]$. The ice cream category, for example, is less important in Store C, but purchases of ice cream single packs are above the store average ($E_c[\Delta_{ks}] = -1.2\%$, $\Delta_{ks} = +69.3\%$). This finding has face validity given the location of Store C; tourists and workers in offices likely prefer single packs over larger ice cream packages. A similar finding can be observed for lunch snacks in that the overall sales across the six retailer categories clearly underestimate the true importance of lunch snacks for shoppers in Store C ($E_c[\Delta_{ks}] = +40.8\%$, $\Delta_{ks} = +200.1\%$). The opposite is true for the barbecue cluster, in which the average category value suggests high importance ($E_c[\Delta_{ks}] = +12.5\%$), but the product cluster reveals that the opposite is true. It is interesting that all values for $E_c[\Delta_{ks}]$ are biased toward zero, most likely because of the aggregation over heterogeneous subgroups of products. This further supports the value of P2V-MAP; the product clusters have the potential to increase the value of descriptive analytics that are common in practice (Wedel and Kannan 2016). A typical category view cannot uncover consumers' preferences, but using P2V-MAP clusters makes this possible.

Discussion and Conclusion

In this article, we have introduced P2V-MAP, a new, exploratory approach for analyzing market structures across all products in a retailer's assortment, using commonly available data. P2V-MAP is based on two recent methodological advances in NLP and machine learning: the SG model and t-SNE. We customize the SG model and its data preprocessing for the application in (grocery) retailing. The SG model scales well to large market basket data, derives meaningful representations of products in latent, linear attribute spaces, and differentiates product similarity and co-occurrence. Next, t-SNE translates the high-dimensional SG model output into a two-dimensional product map, such that similar products in the high-dimensional, attribute space are close to each other on the two-dimensional product map. The use of modern machine learning methods increases P2V-MAP's scalability and makes it possible to update results incrementally. The resulting product maps are easy to use because the latent attributes accurately capture product characteristics that differentiate products. This leads to clearly disjointed map clusters and facilitates the analysis of product complementarity, in addition to product competition.

The empirical application of our approach uses data from a leading, German grocery retailer, and the results regarding market structure, product complementarity and product competition offer good face validity. We verify P2V-MAP's superiority over alternative approaches to mapping market structures in an extensive simulation study and in the empirical application. The simulation study also improves our understanding of how and why P2V-MAP works and validates P2V-MAP in that we show that P2V-MAP results reflect the parameters of the data generating process. We verify that P2V-MAP meets our claims by comparing each of the empirical results to the results of well-established reference models that use loyalty card data to analyze category or product relationships. One unique feature of P2V-MAP is its capability to map products for which no sales data are available (e.g., new products), if product vectors in the latent attribute space can be derived through a linear combination of existing products. In a sample application, we showcase both how retailers can use the P2V-MAP output to analyze the store-specific preferences of their consumers and how the results relate to the assortment depth across clusters.

The results derived through P2V-MAP have relevant managerial implications for retailers. The product clusters on the market structure map can provide information for the design of aisles and shelves. For the retailer studied in the empirical application, we find that some P2V-MAP clusters differ from the retailer's category definition that currently governs aisle and shelf layouts. In these cases, the retailer might consider placing products from different categories next to each other (e.g., organic fruits, vegetables, snacks, bread). Reduced search costs and larger shopping baskets through cross-category selling can increase revenues (Chen, Chen, and Tung 2006). A possible decrease in unplanned spending caused by lower in-store travel distances can be countered by strategic mobile promotions (Hui et al. 2013).

With respect to listing decisions, the product map allows retailers to assess whether their portfolios contain redundant products. In the analysis of the dark chocolate cluster, for example, we find that several Lindt products are nearly identical in the eyes of consumers. A better understanding of a product's uniqueness reduces the risk of cannibalizing sales of other products. Under certain circumstances, P2V-MAP can map products for which no sales data are available; this makes it possible for retailers to anticipate such risks even before listing new products. In addition to reducing inventory costs and utilizing shelf space more efficiently, such changes can make the retailer more attractive, because it carries more relevant items and caters directly to consumers' preferences, which can result in increased store traffic and sales (Briesch, Chintagunta, and Fox 2009; Chernev and Hamilton 2009; Rooderkerk, Van Heerde, and Bijmolt 2013). As a side benefit, the ability to remove products strategically from an assortment should increase retailers' bargaining power in negotiations with manufacturers.

Finally, the insights derived from P2V-MAP have applications in retailer promotions and target marketing. Retailers might consider designing themed leaflets that promote (complementary) products from the same product clusters (e.g., barbecue, vegan, Italian food). In recommender systems, the product map helps to identify suitable alternative offers if particular products are out of stock or removed from the assortment. Promotions informed by the product map can realize cross-selling potential (e.g., organic products, products for children).

This article opens up several opportunities for further research. Our goal with this study was to design and validate P2V-MAP. We provide initial ideas about how the insights generated by P2V-MAP can inform assortment-related decisions. Additional studies should determine the extent to which P2V-MAP improves assortment decisions. For example, its outputs can be used to derive actionable metrics for decision support systems. Linking metrics to outcomes (e.g., in explanatory models) makes it possible to systematically improve marketing actions (Wedel and Kannan 2016). Such approaches should be validated in field experiments (Mantrala et al. 2009).

Another area for future research is the application of P2V-MAP beyond offline grocery retailers. Our approach should work without limitation if the properties of data sets are similar to the characteristics of the data set we used (e.g., data from hardware stores or online grocery retailers). In cases with smaller shopping baskets, the data might require adaptation, such as using a consumer's purchase history instead of a single shopping basket (e.g., Amazon data). A similar approach might be reasonable when analyzing media choices (e.g., Netflix data).

Finally, we hope that our work motivates further methodological research on the application of machine learning techniques to marketing-related problems (compare Sudhir 2016). Extending P2V-MAP (e.g., additional covariates in the SG model) might improve results and widen the method's applicability. In this context, it is worth investigating the extent to which P2V-MAP allows us to track changes in market structure over time and identify trends (Elrod et al. 2002). t-SNE could replace popular dimensionality reduction techniques, such as PCA and MDS, in other marketing applications (e.g., in the analysis of survey data). Combining product embeddings and loyalty card systems presents an opportunity to derive latent attribute vectors for consumers. These can be applied in recommender systems either by computing the similarity between consumers and products or by using similarities between consumers as input for collaborative filtering. The application of embeddings as input layers in deep neural networks has also had compelling success in computer science (LeCun, Bengio, and Hinton 2015); we believe that our work is a promising first step toward deep learning applications in marketing.

Acknowledgments

The authors thank the *JMR* review team for very detailed comments and helpful suggestions. Their input has resulted in additional modifications of the machine learning models, contributed to a better validation of our approach, and led to several analyses in the main text and in the Web Appendix. The authors also thank the employees of the SO1 GmbH, especially Dieter Ebbers, for helpful suggestions regarding the practical application of P2V-MAP.

Associate Editor

Jacob Goldenberg served as associate editor for this article.

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: Daniel Guhl gratefully acknowledges financial support by the Deutsche Forschungsgemeinschaft (DFG) through CRC TRR 190.

References

Abadi, Martin, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, et al. (2015), "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," software, tensorflow.org.

Anderson, Annie S. and Susan E. Morris (2000), "Changing Fortunes: Changing Food Choices," *Nutrition & Food Science*, 30 (1), 12–15.

Asgari, Ehsaneddin and Mohammad M.K. Mofrad (2015), "Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics," *PloS One*, 10 (11), e0141287.

Blattberg, Robert C., Byung-Do Kim, and Scott A. Neslin (2008), *Database Marketing: Analyzing and Managing Customers*. Berlin: Springer.

Bradlow, Eric T., Manish Gangwar, Praveen Kopalle, and Sudhir Voleti (2017), "The Role of Big Data and Predictive Analytics in Retailing," *Journal of Retailing*, 93 (1), 79–95.

Breugelmans, Els, Katia Campo, and Els Gijsbrechts (2007), "Shelf Sequence and Proximity Effects on Online Grocery Choices," *Marketing Letters*, 18 (1/2), 117–33.

Briesch, Richard A., Pradeep K. Chintagunta, and Edward J. Fox (2009), "How Does Assortment Affect Grocery Store Choice?" *Journal of Marketing Research*, 46 (2), 176–89.

Chen, Yen-Liang, Jen-Ming Chen, and Ching-Wen Tung (2006), "A Data Mining Approach for Retail Knowledge Discovery with Consideration of the Effect of Shelf-Space Adjacency on Sales," *Decision Support Systems*, 42 (3), 1503–20.

Chernev, Alexander and Ryan Hamilton (2009), "Assortment Size and Option Attractiveness in Consumer Choice Among Retailers," *Journal of Marketing Research*, 46 (3), 410–20.

Chintagunta, Pradeep K. (1992a), "Estimating a Multinomial Probit Model of Brand Choice Using the Method of Simulated Moments," *Marketing Science*, 11 (4), 386–407.

Chintagunta, Pradeep K. (1992b), "Investigating Heterogeneity in Nested Logit Models: A Modeling Approach and Empirical Analysis," *International Journal of Research in Marketing*, 9 (2), 161–75.

DeSarbo, Wayne S. and Rajdeep Grewal (2007), "An Alternative Efficient Representation of Demand-based Competitive Asymmetry," *Strategic Management Journal*, 28 (7), 755–66.

Elrod, Terry and Michael P. Keane (1995), "A Factor Analytic Probit Model for Representing Market Structure in Panel Data," *Journal of Marketing Research*, 32 (1), 1–16.

Elrod, Terry, Gary J. Russell, Allan D. Shocker, Rick L. Andrews, Lynd Bacon, Barry L. Bayus, et al. (2002), "Inferring Market Structure from Customer Response to Competing and Complementary Products," *Marketing Letters*, 13 (3), 221–32.

Erdem, Tülin (1996), "A Dynamic Analysis of Market Structure Based on Panel Data," *Marketing Science*, 15 (4), 359–78.

Fader, Peter S. and Bruce G.S. Hardie (1996), "Modeling Consumer Choice Among SKUs," *Journal of Marketing Research*, 33 (4), 442–52.

Gower, John C. (1966), "Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis," *Biometrika*, 53 (3/4), 325–38.

Hertz, John A., Anders S. Krogh, and Richard G. Palmer (1991), *Introduction to the Theory of Neural Computation*. Boston: Addison-Wesley/Addison Wesley Longman.

Hruschka, Harald (2014), "Analyzing Market Baskets by Restricted Boltzmann Machines," *OR Spectrum*, 36 (1), 209–28.

Hui, Sam K., J. Jeffrey Inman, Yanilu Huang, and Jacob Suher (2013), "The Effect of In-Store Travel Distance on Unplanned Spending: Applications to Mobile Promotion Strategies," *Journal of Marketing*, 77 (2), 1–16.

Jacobs, Bruno J.D., Bas Donkers, and Dennis Fok (2016), "Model-Based Purchase Predictions for Large Assortments," *Marketing Science*, 35 (3), 389–404.

Jarvis, Wade and Steve Goodman (2005), "Effective Marketing of Small Brands: Niche Positions, Attribute Loyalty and Direct Marketing," *Journal of Product & Brand Management*, 14 (5), 292–99.

Kim, Jun B., Paulo Albuquerque, and Bart J. Bronnenberg (2011), "Mapping Online Consumer Search," *Journal of Marketing Research*, 48 (1), 13–27.

Kingma, Diederik P. and Jimmy Ba (2014), "Adam: A Method for Stochastic Optimization," arXiv preprint arXiv:1412.6980.

Kök, A. Gürhan, Marshall L. Fisher, and Rammath Vaidyanathan (2015), "Assortment Planning: Review of Literature and Industry Practice," in *Retail Supply Chain Management*, Chapter 6, Narendra Agrawal, and Stephen A. Smith, eds. Boston: Springer, 175–236.

Kullback, Solomon and Richard A. Leibler (1951), "On Information and Sufficiency," *The Annals of Mathematical Statistics*, 22 (1), 79–86.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015), "Deep Learning," *Nature*, 521 (7553), 436–44.

Lee, Thomas Y. and Eric T. Bradlow (2011), "Automated Marketing Research Using Online Customer Reviews," *Journal of Marketing Research*, 48 (5), 881–94.

Manchanda, Puneet, Asim Ansari, and Sunil Gupta (1999), "The 'Shopping Basket': A Model for Multicategory Purchase Incidence Decisions," *Marketing Science*, 18 (2), 95–114.

Mantrala, Murali K., Michael Levy, Barbara Kahn, Edward J. Fox, Peter Gaidarev, Bill Dankworth, and Denish Shah (2009), "Why Is Assortment Planning So Difficult for Retailers? A Framework and Research Agenda," *Journal of Retailing*, 85 (1), 71–83.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean (2013), "Distributed Representations of Words and Phrases and Their Compositionality," in *Advances in Neural Information Processing Systems* 26, C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, eds. San Diego: Neural Information Processing Systems Foundation, Inc, 3111–19.

Mitra, Bhaskar, Eric Nalisnick, Nick Craswell, and Rich Caruana (2016), "A Dual Embedding Space Model for Document Ranking," arXiv preprint arXiv:1602.01137.

Neslin, Scott A. and Harald J. van Heerde (2009), "Promotion Dynamics," *Foundation and Trends in Marketing*, 3 (4), 177–268.

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.

Olsson, Ulf (1979), "Maximum Likelihood Estimation of the Polychoric Correlation Coefficient," *Psychometrika*, 44 (4), 443–60.

Padel, Susanne and Carolyn Foster (2005), "Exploring the Gap Between Attitudes and Behaviour: Understanding Why Consumers Buy or Do Not Buy Organic Food," *British Food Journal*, 107 (8), 606–25.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, et al. (2011), "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research*, 12 (October), 2825–30.

Phi, Van-Thuy, Liu Chen, and Yu Hirate (2016), "Distributed Representation-Based Recommender Systems in E-Commerce," *DEIM Forum*, (accessed March 9, 2019), http://db-event.jpn.org/deim2016/papers/27.pdf.

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.

Rooderkerk, Robert P., Harald J. van Heerde, and Tammo H. Bijmolt (2013), "Optimizing Retail Assortments," *Marketing Science*, 32 (5), 699–715.

Rousseeuw, Peter J. (1987), "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis," *Journal of Computational and Applied Mathematics*, 20, 53–65.

Russell, Gary J. and Ann Petersen (2000), "Analysis of Cross Category Dependence in Market Basket Selection," *Journal of Retailing*, 76 (3), 367–92.

Stassen, Robert E., John D. Mittelstaedt, and Robert A. Mittelstaedt (1999), "Assortment Overlap: Its Effect on Shopping Patterns in a Retail Market When the Distributions of Prices and Goods Are Known," *Journal of Retailing*, 75 (3), 371–86.

Sudhir, K. (2016), "Editorial—The Exploration-Exploitation Tradeoff and Efficiency in Knowledge Production," *Marketing Science*, 35 (1), 1–9.

Train, Kenneth E. (2009), *Discrete Choice Methods with Simulation*, 2nd ed. Cambridge, UK: Cambridge University Press.

Van der Maaten, Laurens (2014), "Accelerating t-SNE Using Tree-Based Algorithms," *Journal of Machine Learning Research*, 15 (October), 3221–45.

Van der Maaten, Laurens and Geoffrey Hinton (2008), "Visualizing Data Using t-SNE," *Journal of Machine Learning Research*, 9 (November), 2579–605.

Van Nierop, Erjen, Dennis Fok, and Philip Hans Franses (2008), "Interaction Between Shelf Layout and Marketing Effectiveness and Its Impact on Optimizing Shelf Arrangements," *Marketing Science*, 27 (6), 1065–82.

Venkatesan, Rajkumar and Paul W. Farris (2012), "Measuring and Managing Returns From Retailer-Customized Coupon Campaigns," *Journal of Marketing*, 76 (1), 76–94.

Vinh, Nguyen X., Julien Epps, and James Bailey (2010), "Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance," *Journal of Machine Learning Research*, 11 (October), 2837–54.

Vylomova, Ekaterina, Laura Rimell, Trevor Cohn, and Timothy Baldwin (2015), "Take and Took, Gaggle and Goose, Book and Read: Evaluating the Utility of Vector Differences or Lexical Relation Learning," arXiv:1509.01692.

Walmart (2005), "Our Retail Divisions," (accessed December 30, 2016), http://corporate.walmart.com/_news_/news-archive/2005/01/07/our-retail-divisions.

Walmart (2016), "Walmart.com's History and Mission," (accessed December 30, 2016), https://help.walmart.com/app/answers/detail/a_id/6.

Wedel, Michel and P.K. Kannan (2016), "Marketing Analytics for Data-Rich Environments," *Journal of Marketing*, 80 (6), 97–121.