# Design Science

# Predicting product co-consideration and market competitions for technology-driven product design: a network-based approach

Mingxian Wang[1], Zhenghui Sha[2], Yun Huang[3], Noshir Contractor[3], Yan Fu[1] and Wei Chen[4]

1 *Global Data, Insight and Analytics, Ford Motor Company, Dearborn, MI, USA*
2 *System Integration and Design Informatics Laboratory, University of Arkansas, Fayetteville, AR, USA*
3 *Science of Networks in Communities, Northwestern University, Evanston, IL, USA*
4 *Integrated Design Automation Laboratory, Northwestern University, Evanston, IL, USA*

## Abstract

We propose a data-driven network-based approach to understand the interactions among technologies, products, and customers. Specifically, the approach enables both a qualitative understanding and a quantitative assessment of the impact of technological changes on customers' co-consideration behaviors (decision of cross-shopping) and as a consequence the product competitions. The uniqueness of the proposed approach is its capability of predicting complex co-consideration relations of products as a network where both descriptive analyses (e.g., network statistics and joint correspondence analysis) and predictive models (e.g., multiple regressions quadratic assignment procedure) are employed. The integrated network analysis approach features three advantages: (1) It provides an effective visual representation of the underlying market structures; (2) It facilitates the evaluation of the correlation between customers' consideration preferences and product attributes as well as customer demographics; (3) It enables the prediction of market competitions in response to potential technological changes. This paper demonstrates the proposed network-based approach in a vehicle design context. We investigate the impacts of the fuel economy-boosting technologies and the turbocharged engine technology on individual automakers as well as the entire auto industry. The case study provides vehicle engineers with insights into the change of market competitions brought by technological developments and thereby supports attribute decision-making in vehicle design.

**Key words:** network analysis, data-driven design, consideration preference, customer preference, correspondence analysis, market competition, technology development

the **Design Society**
*a worldwide community*

**CAMBRIDGE**
UNIVERSITY PRESS

## 1. Introduction

Forecasting customers' responses and market impacts brought by the changes to design attributes/features is essential before launching major technological changes on future generations of products. To prioritize product attributes and improve competitive advantages, an in-depth investigation and assessment of technological opportunities, consumer needs, product positions, and competitive
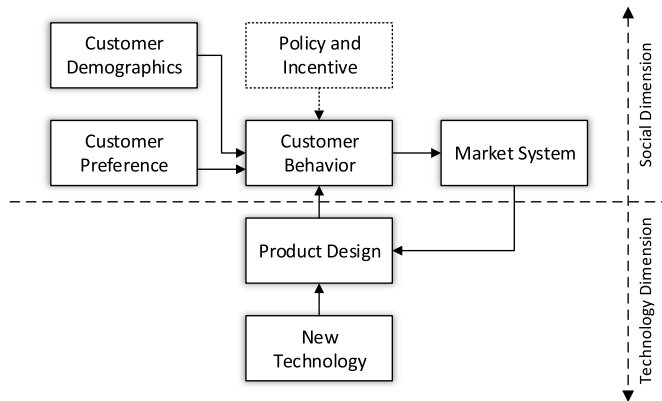
**Figure 1.** A social–technical system for understanding the interactions among technologies, products, customers and the market.

landscapes are necessary in the phases of product design, strategic planning, and product development.

This study explores the intersection between traditional engineering design research and product marketing research, in an attempt to analyze the impact of introducing new technologies on market competition through a data-driven consumer preference model. We consider the complex relationship among technologies, products, and customers as a socio-technical system. As shown in Figure 1, customer behavior, reflected in their consideration and choice decisions, are affected by product design, represented by product attributes (such as color and size), customer demographics (such as income, social status, and education), and customers' preferences. Besides incentives and policies, a common way to promote product market share is to upgrade product attributes (e.g., by utilizing new technologies) based on new market trends and customer needs. Following this interacting relationship, this work offers a new network-based customer preference modeling technique by modeling an interacting system constituted by technologies, products, customers and the market.

Among different types of customers' behaviors, we are particularly interested in understanding customer *consideration* behavior. Co-consideration describes the situation where a customer concurrently considers multiple products in cross-shopping activities. A customer will not purchase a product if the product is not considered. The consideration behavior involves comparison and evaluation of product alternatives and thus is a crucial step before making the final choice. Co-consideration also implies the market competition between products or brands, which is crucial to companies' product launching and positioning plans as well as market strategies. Existing studies (Hauser & Wernerfelt 1990; Shocker 1991; Hauser 2014) have shown that customers' consideration sets are small, frequently ranging from two to six options due to humans' limited capacity in processing decision-related information. The questions of how to characterize customers' co-consideration decisions as well as how to evaluate the influence of product attributes on those decisions become valuable yet daunting, mainly due to the unobserved preferences of customers and their heterogeneous decision-making behaviors. Consideration sets generally describe the market competitions

from the eyes of customers. Successful modeling of product co-consideration relations helps understand the embedded market competitions and provides new opportunities to enterprises to formulate design solutions, address customer needs, and make strategic moves. Therefore, a quantitative modeling framework that enables the analysis of interaction among customer consideration decisions, product configurations, and market complexities is indispensable in product design research.

Our **research objective** is twofold: (1) integrating descriptive multivariate analysis with predictive network models to investigate product co-consideration relations as well as the potential driving factors related to customer demographics and product attributes; (2) characterizing how customer consideration preference and technology development change the competition map in a market. To achieve the objective, we take a data-driven network-based approach, where we model products as nodes and the co-consideration relations between any two products as links. With such a construct, the key idea is to understand and predict co-consideration relations as a function of explanatory networks derived from the associations of product design and customer demographical attributes.

The remainder of the paper is structured as follows. Section 2 reviews relevant studies on technology acceptance modeling, social behavior modeling, consumer preference modeling, association analysis, and machine learnings. To address the limitation of existing approaches, we present our data-driven network-based approach in Section 3 in a step-by-step format to guide and facilitate the implementation. In Section 4, our approach is applied to understand vehicle's co-consideration relations in China market using survey data on new vehicle buyers. As a case study, we predict how the technological changes, including fuel economy-boosting techniques and the turbocharged engine, would affect competitions between vehicle lines, brands, and the whole industry. Section 5 concludes the paper with additional insights and closing thoughts.

## 2. Literature review

Understanding the impact of new technologies and predicting the preference of customers have drawn continuous interests in the engineering design community since the 1980s. The technology acceptance model (TAM) proposed by Fred Davis (Davis 1986) argues that the use of technology is a system response that can be explained and predicted by customers' motivation, which, in turn, directly influence system's features. Such inner relationships correspond to the interactions among products, customers, and the market system as shown in Figure 1. Davis suggests explaining users' motivation by three factors: *Perceived Ease of Use*, *Perceived Usefulness*, and *Attitudes Toward Using* the technology. Later development of TAM has evolved to many versions by substituting *Attitudes Toward Using* with *Behavioral Intention* (Davis, Bagozzi & Warshaw 1989), by adding extra variables as antecedents to *Perceived Usefulness* variable, referred as TAM2 (Venkatesh & Davis 2000), and by identifying the antecedents to *Perceived Ease of Use* variable, referred as TAM3 (Venkatesh & Bala 2008). Later research has also extended the TAM by including the social influence due to the increasing popularity of the Internet and social media (Malhotra & Galletta 1999; Venkatesh & Morris 2000; Hsu & Lu 2004). Among extensive studies in this field, most work is empirical and qualitative. Experimental data is primarily used to test hypothesized causal relationships. From the methodological perspective, the main limitation is

the employment of self-reported use data (e.g., a subjective measurement of verbal description) instead of actual use data. This would lead to weak validation and biased results by respondents (Lee, Kozar and Larsen 2003).

Relating to market competition, existing literature has used game-theoretic models to understand the impact of technology quantitatively. Thatcher (2004) developed a two-stage model of duopoly competition to examine the impact of information technology (IT) investments in product design tools on improving product quality and price, firm productivity and profits, and consumer welfare. Thatcher observed that profit-maximizing firms often leverage technology-based design tools to improve product quality, firm profits, and consumer welfare, but at the expense of productivity. Other theoretical studies, such as (Thatcher & Oliver 2001), are similar but the analysis is performed in the monopolist context. Since the early 1970s, many studies have investigated the impact of technology, especially in the field of IT on different levels and categories of systems, including economic, industrial, social, etc. However, most studies do not address the impact at the level of individual customers. Even though some studies (Brynjolfsson 1996; Hitt & Brynjolfsson 1996; Thatcher 2004) have realized the importance of individual customers, they mainly provided economic insights, such as the impact on consumer welfare and surplus, instead of design perspectives.

In creating analytical customer preference models, discrete choice analysis (DCA) has been widely adopted to understand how customers make trade-offs on product attributes (Wassenaar & Chen 2003; Train 2009). Rooted in econometrics, DCA, in essence, uses utilities to evaluate the best option from a consideration set. DCA can predict choice probabilities, e.g., the probability of buying a Toyota Prius, given the consideration set of each customer and the constructed utility function; however, it cannot capture the complex association relations of alternatives in customers' consideration set (Wang *et al.* 2016). As an extension to DCA, recent studies proposed heuristics-based models for characterizing consideration decisions (Gilbride & Allenby 2004; Cantillo & de Dios Ortúzar 2005; Gaskin *et al.* 2007; Dieckmann, Dippold & Dietrich 2009; Hauser *et al.* 2010; Morrow, Long & MacDonald 2014). However, existing work in engineering design only explored the suitability of various forms of non-compensatory and compensatory utility models using synthetic data generated by pre-defined adjunctive rules (Morrow *et al.* 2014). These methods do not directly address three critical questions: (1) what products tend to be in the same consideration set, (2) what customer and product attributes explain the product co-consideration, and (3) how the similarity (or dissimilarity) of products and customers affect product co-consideration and market competition.

Compared to DCA-based econometric models, our network-based approach has two advantages. First, the proposed co-consideration relations can be described and depicted with simple graphical structures based on customer cross-shopping data. Our method focuses more on the relations of data instead of treating products as separated, individual entities. This feature more closely resembles the interdependent cross-shopping decisions. Second, decisive factors associated with product co-considerations and market competitions can be identified and evaluated through quantitative network modeling. In addition to the impact of product and customer attributes studied in the traditional econometric analysis, network analysis can also evaluate the impact of relations between product attributes (e.g., similarities, differences) on customers' behaviors.

Recent years have seen the use of machine learning techniques for predicting customers' preferences. One of the related techniques to our study is the market basket analysis (Mostafa 2015), which uses association rules to discover what products (e.g., beer and dippers) are bought together. Recent effort has been made to integrate the market basket analysis with network analysis to uncover patterns beyond relations between two products (Raeder & Chawla 2011; Kim, Kim & Chen 2012). However, existing analyses disclose the association relations but are not aimed for explaining why products are associated (e.g., co-considered). In our research, network modeling approach is used to expand descriptive analyses to model-based predictive approaches to understand consumer behaviors and assess quantitatively the impact of underlying product design attributes. Our models are built from the perspectives of designers with the goal of understanding customers' preferences in designing consumer-desired products.

Our proposed approach also falls into the broader category of the recommender systems/engines, which attempt to predict future likes and interests by mining data on past user activities (Resnick & Varian 1997). Although recent work on recommender systems has also utilized network representation for input data (Zhou *et al.* 2007; Fiasconaro *et al.* 2015; Yu *et al.* 2016), many differences arise in the process of modeling for an explanatory versus a predictive interest. Most existing recommendation algorithms are inspired and progressed from the classical physical processes, such as random walk, heat diffusion, and preferential diffusion (Yu *et al.* 2016). The focus of these methods is more on recommendation (prediction) with only a slight and indirect relation to casual explanation. Our proposed network models, in comparison, possess explanatory power. The inference capabilities inherited in our model offer the conveniences to describe statistically the impact of product attributes on the market behaviors. Again while traditional recommender systems offer suggestions for customers' decision-making in product purchase, our models are used to assist designers' decision-making for improving product attributes.

In addition, our work has benefited from the system thinking by treating the complex socio-technical system as a single entity. Network models are devised to fill the gap between separate domains of technology innovations, market competitions, and customer preference. This system's view answers how the separate domains link and affect each other, for example, how technology-driven changes affect product co-considerations, how customers' consideration decisions impact product competitions, and how to use market competition structures to guide the product design and development. In a market with diverse products, the lack of such knowledge would impede enterprises' strategic moves in adopting emerging technologies.

## 3. Proposed network-based approach

### 3.1. Background and related works

A novel network-based framework has recently been developed to understand the complex relations between customers and products (Wang *et al.* 2016). This network-based approach applies network analysis and graph modeling to study relational patterns, structural features, and evolving relationships among system components. Advantages of network analysis have been widely recognized by social science researchers. The problems involving interdependent

behaviors and complex patterns are explained by combining multiple generative mechanisms such as self-interest, collective action, social exchange, balance, homophily, contagion, and co-evolution (Monge & Contractor 2003). As an effort to promote network analysis for design problems, our previous research introduced product networks based on co-consideration data and proposed a heuristic algorithm to predict missing choice sets in supporting choice modeling (Wang & Chen 2015). Later, the unidimensional product network is extended to a multidimensional customer–product network (MCPN) (Wang *et al.* 2016). The structure is developed to model multiple types of relations, including co-consideration decisions, choice decisions, social interactions, and product interdependence. Built upon the MCPN framework presented in our earlier work (Wang *et al.* 2016), for the first time, this paper presents the use of both descriptive (graphical) network analysis approach and a quantitative inferential modeling approach to examine the customer preferences through the product co-consideration network. Besides, we demonstrate how the two methods are integrated, complementing each other to inform design decisions. This paper also contributes to developing use cases of network approaches in design research, with examples of studying the market impact of technological innovations.
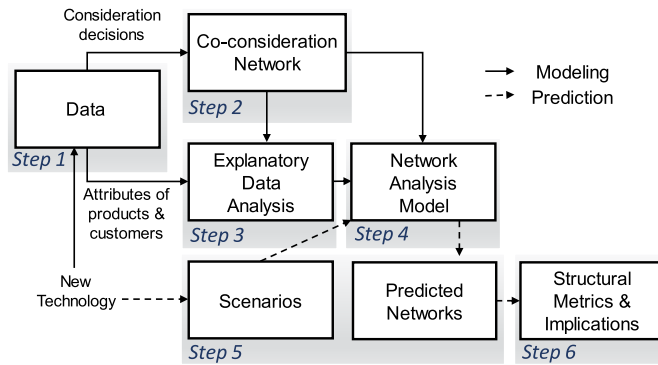
## 3.2. Overview of the network-based approach

As shown in Figure 2, our data-driven, network-based approach models a market as a single network entity, which captures the interdependence among product relations driven by customer preferences. The two central elements of the approach are the qualitative exploration tool (Step 3) and the quantitative analytical model (Step 4). The two elements together provide the statistical modeling of co-consideration behaviors with respect to the various potential drivers, such as product design attributes, customer demographic attributes, and customer perceived product characteristics, as well as the similarity or difference of product profiles and market positions. For example, the homophily effect (McPherson, Smith-Lovin & Cook 2001) represents the extent to which products form links with similar versus dissimilar alternatives. Examining this effect can uncover why products are co-considered and compared against each other. As to the forecast of technological impacts, the topological structures of predicted networks under new technological scenarios are analyzed, from which insights on market competitions and product association can be obtained accordingly.

This work considers market competition from both the perspective of a single product and the perspective of producers (i.e., a group of products within the same brand). From the next section onwards, we present each step of the proposed approach following the flow diagram in Figure 2.

## 3.3. Step 1 – Data collection

The first step involves data collection and preparation. The proposed approach requires data to cover at least two aspects: (1) competing alternatives that customers considered, and (2) design attributes on alternatives. For the former, various formats of data such as surveys, web-browsing logs, data from comparison engines can be used to identify consideration decisions made by individual customers. For the latter, product specifications or customer-stated product configurations could be used for the description of product profiles. Beyond the

**Figure 2.** Overview of the proposed approach.

two requirements, additional information on customer demographics and their preferences could improve model performance.

### 3.4. Step 2 – Characterizing product association as a co-consideration network

The network construction starts in Step 2 with the data collected in the previous step. The goal is to characterize customers' consideration preferences. We define each node as a representation of a product. A link exists between two nodes if products are co-considered and compared by the majority of customers. In other words, the links between products reflect the proximity or similarity of two products in customers' considerations. For example, given that many customers consider 'Ford Edge', 'Ford Changan Kuga', and 'Honda Dongfeng CR-V' together, we may extract the three vehicle models and establish links between each pair of them (see Figure 3). The strength of the links can be evaluated by standard metrics of association rules, showing how likely the two products are co-considered by customers. In Figure 3, the link strength between Edge and CR-V (1.2) is smaller than the strength between Kuga and CR-V (2.4), implying that CR-V is more likely to be considered with Kuga together than with Edge.

In the network shown in Figure 3, the size of a node is proportional to the degree centrality of the node (i.e., the number of links), which can be used to study the range of consideration on a product and its impact on market share (Wang *et al.* 2015). The node colors highlight the clusters of nodes (Clauset, Newman & Moore 2004), which represent the groups of vehicle models frequently considered together by many customers. Details of network constructions, link strength calculations, and community detections are discussed in Raeder & Chawla (2011) and Wang *et al.* (2016). Based on the constructed network, in the next step, we introduce a tool for exploratory analysis.

### 3.5. Step 3 – Understanding key influencing factors to co-consideration decisions

In Step 3, we use correspondence analysis, and in particular, joint correspondence analysis (JCA) (Benzécri 1973; Greenacre 2007) as a descriptive analysis tool to identify the underlying key drivers among product and customer attributes to
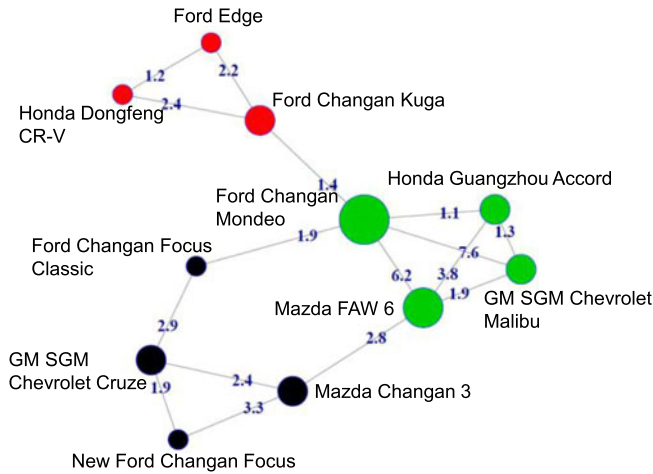
**Figure 3.** Illustrative network of vehicle co-considerations.

the formation of a product co-consideration network. The multivariate nature of such an approach facilitates the detection of relationships among the attribute categories (i.e., attribute levels of customers and products) and objects (i.e., products and customers). The correspondence analysis technique maximizes the interrelationships between the rows and columns of a multiway data table for dimensional reduction. Orthogonal components and factor scores created on the levels of categorical attributes allow the construction of visual plots whose structures can be easily interpreted (Greenacre 2007). In this work, the generated visual plots embed the community structures of a product co-consideration network as well as the relations among customer and product attributes of interests.

We use conventional notations (Greenacre & Blasius 2006) to demonstrate the JCA approach. Assume there are $N$ customers and $v_1, \ldots, v_q$ categorical attributes such as vehicle models and income levels. Each $v_j$ has $L_j$ levels $1, \ldots, L_j$, and a binary indicator matrix $\mathbf{Z}^{(j)}$ with $N \times L_j$ dimensions indicates the associations between customers and attribute $v_j$. $\mathbf{Z}_{il}^{(j)} = 1$ if and only if $v_j^{(i)} = l$, i.e., the attribute $j$ of customer $i$ has the level $l$. All $\mathbf{Z}^{(j)}$ can then be horizontally concatenated to form a large indicator matrix $\mathbf{Z}$ of $N \times J$, where $J = L_1 + \cdots + L_q$ is the total number of categorical levels for all input variables $x$.

For example, Table 1 shows an indicator matrix $\mathbf{Z}$ with five customers as row entries and two categorical attributes as column entries: vehicle model $v_1$ lists all vehicle alternatives customers could consider, and income level $v_2$ divides all customers based on three levels of incomes. Because the indicator matrix $\mathbf{Z}$ can take up considerable memory space when the number of rows (i.e., customers) is large, JCA typically operates on a condensed Burt matrix $\mathbf{B} = \mathbf{Z}'\mathbf{Z}$, which equals the cross-tabulation of all categorical levels. Given the Burt matrix $\mathbf{B}$, the coordinates of columns with respect to the principal axes can be computed by singular value decomposition (SVD) (Greenacre & Blasius 2006), and the inertia can be obtained by iteratively updating the solution (Greenacre 2007). The Appendix includes more technical details about the approach.
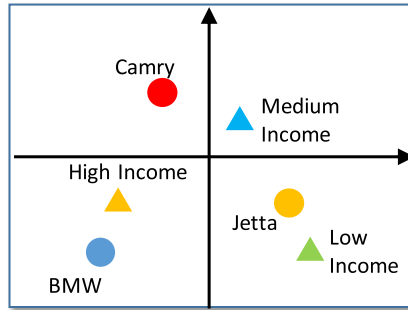
**Figure 4.** Demonstrative perceptual map generated by JCA. Vehicle models are shown in dots and income levels in triangles. Two vehicles are close to each other if they are considered by the same customers; two income levels are close to each other if they are tied to the same vehicle buyer; a vehicle model and an income level are placed close to each other if customers considering the vehicle often have such an income level.

**Table 1.** Demonstrative indicator matrix in joint correspondence analysis, with customers as row entries, and vehicle models and income levels as column entries.

|  | Vehicle model $v_1$ | | | Income level $v_2$ | | |
|---|---|---|---|---|---|---|
|  | Jetta | Camry | BMW7 | Low | Mid | High |
| Customer 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| Customer 2 | 1 | 0 | 0 | 0 | 1 | 0 |
| Customer 3 | 1 | 0 | 0 | 1 | 0 | 0 |
| Customer 4 | 0 | 1 | 0 | 0 | 1 | 0 |
| Customer 5 | 0 | 1 | 0 | 0 | 0 | 0 |

Being a descriptive technique, JCA emphasizes on the graphical representation of relational data. The visual output of JCA draws the first two dimensions of the principal coordinates of columns jointly in a Euclidean 2D space. As noted in Figure 4, the demonstrative output simultaneously displays the examined vehicle models in dots and the income levels in triangles. The distances between two points in the space represent the strengths of correlations between a pair of variables. For example, according to Figure 4, BMW attracts more high-income customers given the closeness of the two entities; while Jetta is more attainable for low-income customers. The axes represent the principal factors that differentiate vehicle models from one to another as the analysis result of defined relations in Table 1. A more concrete example can be found in Section 4.3. By analyzing the visual plot, designers can identify the major factors and factor levels that drive certain vehicle models to be co-considered, and further, eliminate irrelevant or redundant attributes before constructing more advanced predictive network models.
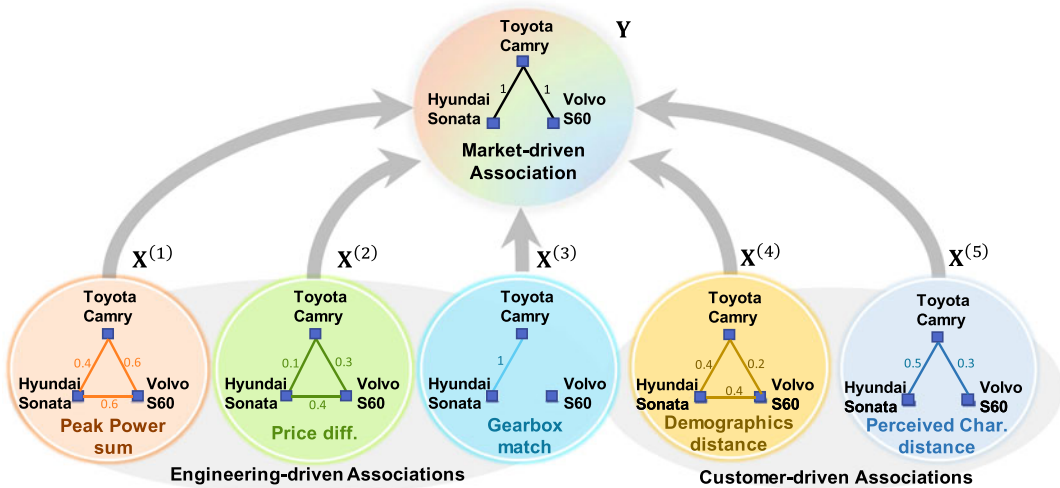
**Figure 5.** Illustration of MRQAP Model. Co-consideration decisions (**Y** at top) are predicted using engineering-driven associations and customer-driven associations created by attribute data (**X**s at bottom).

## 3.6. Step 4 – A network analysis model to predict co-consideration decisions

Built on the descriptive analysis, the goal of Step 4 is to develop a predictive model and further analyze the network patterns. This paper utilizes the multiple regression quadratic assignment procedure (MRQAP) (Krackhardt 1988), which is designed to predict the presence of a relation by explanatory networks formed along various dimensions (Krackhardt 1988), such as product design attributes and customer demographics.

As illustrated in Figure 5, the idea of the MRQAP approach is to decompose the complex co-consideration relations (i.e., market-driven associations) into a function of networks that represent similarities or differences of product configurations (i.e., engineering-driven associations) and customer demographics (i.e., customer-driven associations). The coefficients identified in an MRQAP model indicate the impacts of individual effect networks on forming co-consideration relations. The response **Y** is an adjacency matrix of the co-consideration network, where each weighted link indicates how much dependence two products have. At the bottom, product attributes are vectorized as effect networks, represented by $\mathbf{X}^{(k)}$. Each of them captures the associations between pairs of products based on various arithmetic operations of attributes, for example, the sum of peak powers, the difference in prices, and the match of fuel type shown in Figure 5. The unique aspect of MRQAP is to use simple product networks **X** (created using attribute data) to *predict* the structure of the observed complex decision network **Y** (created using co-consideration data). Let $Y_{ij}$ be the dependent co-consideration link between vehicles $i$ and $j$, and $X_{ij}^{(k)}$ be the $k^{\text{th}}$ covariate (attribute) for the same link observation. The MRQAP model is analogous to the standard logistic regression element-wise on network connections, where the systematic component is given in Eqn. (1). In estimating the $p$-values, permutation techniques are often used to handle correlated observations in network data. Examples include DSP and FLSP

10/34

approaches documented in (Dekker, Krackhardt & Snijders 2007).

$$Pr(Y_{ij} = 1) = E(Y_{ij}) = \frac{\exp(\beta_0 + \beta_1 X_{ij}^{(1)} + \cdots + \beta_n X_{ij}^{(n)})}{1 + \exp(\beta_0 + \beta_1 X_{ij}^{(1)} + \cdots + \beta_n X_{ij}^{(n)})}. \quad (1)$$

As shown in Table 2, the explanatory effect networks in our MRQAP model allow the modeling of two types of effects: the attribute-based main effects and the homophily effects. The attribute-based main effects examine whether products with a specific attribute (or with a high-valued attribute) are more likely to have consideration links than products without the attribute (or with a low-valued attribute). An example is the peak power sum. A positive parameter $\beta$ associated with this effect network indicates that vehicles with a higher sum of powers are more likely to be co-considered compared to vehicles with a lower sum of powers. Thus, a vehicle with a high peak power tends to express more co-consideration links. The homophily effects, originated from the social network literature, represent the tendency of entities to associate and bond with similar others. In the context of the product co-consideration network, the homophily effects examine whether products with similar attributes are more likely to have consideration links with each other. One example is the price difference. A negative parameter $\beta$ suggests that vehicles with a smaller difference in price are more likely to be considered together. In the literature, it is often desirable to have both main effects and homophily effects included in one network model. The main effects control the direct impacts associated with the levels of attributes, whereas the homophily effects explain the co-consideration decisions by the similarities or differences of products in terms of their attributes. Table 2 presents the guidelines for creating explanatory networks in MRQAP for different types of attributes such as binary, categorical, and continuous. For the product attributes under (a)–(c), the strength of the tie $X_{ij}$ in an explanatory network is determined by the corresponding attributes $x_i$ and $x_j$ associated with the linked products. Beyond product attributes, we may also introduce non-product related attributes (d). For example, the customer demographics can be included in the model to allow the prediction of technological impacts in a completely new market following different customer profiles. The multivariable association techniques, such as JCA coordinates generated from Step 3, can be used to express the similarities of the non-product related attributes as the distances of product points ($x_i$ and $x_j$) in a geometric space.

It is noted that the logistic form of MRQAP model implicitly binarizes the co-consideration network **Y**. Our dichotomization is equivalent to the one-vs-rest strategy in the binary transformation. We put the not-co-considered links (more than 91% cases) in one class and the rest in another class. Therefore, we try to reveal the underlying mechanisms of customers identifying product alternatives. It is true that modeling a binary network, while computationally simpler, is not as rich as the weighted network. We choose the binarization setting, because our underlying lift samples are non-randomly selected. To correct the sampling bias, a two-stage estimation method is often needed (Heckman 1976). This work fits into the first stage model of the two-stage correction, where the focus is on predicting whether two products are co-considered or not. The extent of how often two products are co-considered (reflecting the intensity of competition) can be observed by modeling networks with increased cutoff values or through the construction of more complicated models.

**Table 2.** Constructing explanatory networks of attributes.

| Configuration | Statistic | | Network effect |
|---|---|---|---|
| **(a) Binary product attributes** | | | |
| Sum network | $X_{ij} = x_i + x_j$ | (2) | Attribute-based main effect |
| Match network | $X_{ij} = I\{x_i = x_j\}$ | (3) | Homophily effect |
| (b) **Categorical product attributes** | | | |
| Match network | Same as (3) | | Homophily effect |
| (c) **Continuous product attributes (standardized)** | | | |
| Sum network | Same as (2) | | Attribute-based main effect |
| Difference network | $X_{ij} = |x_i - x_j|$ | (4) | Homophily effect |
| **(d) Non- product related attributes** | | | |
| Distance network | $X_{ij} = \|x_i - x_j\|_2$ | (5) | Homophily effect |

Note:

- $I\{\cdot\}$ represents the indicator function.
- $|\cdot|$ represents the absolute-value norm on the 1-dim space.
- $\|\cdot\|_2$ represents the $L_2$-norm on the n-dim Euclidean space.

### 3.7. Step 5 – Network prediction under technological change scenarios

In Step 5, the MRQAP model obtained from Step 4 is used to predict the co-consideration relations between products under specific scenarios of technological changes. The influence induced by new technologies on market competition structures (measured by the topology of a co-consideration network) can then be studied by mapping the technological changes to the change of corresponding product design attributes. Specifically, with the change of the effect networks **X**, co-consideration relations in the form of product co-consideration network **Y** can be predicted by (1).
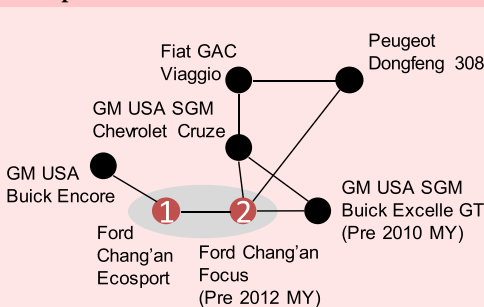
### 3.8. Step 6 – Evaluation of network structure metrics

Finally, in Step 6, the change of the network topology is characterized and quantified by various network metrics to provide insights for marketing strategies and design decisions. In this paper, a set of network metrics, as shown in Table 3, are developed exclusively for the co-consideration network. Using vehicle as an example, a set of metrics are used to characterize the global network (market-wide competition) such as $\bar{d}, c_G, \bar{c}$; while the remaining metrics $N_c, d, \bar{d}, d', \bar{d}', c, \bar{c}$ and $b$ are used to characterize the local network centered around a product node or a set of product nodes that belong to the same brand. This set of metrics represents our pilot effort toward guiding and inspiring model practitioners in design-related activities. There may be other network structural features that worth looking into. However, given the scope of this paper, our discussion is limited to the below selected metrics.

**Table 3.** Examples of network metrics used to quantify the properties of a co-consideration network.

| Network metrics | Definition | Interpretation in vehicle co-considerations and market competitions |
|---|---|---|
| Degree ($d$) | The number of links of a node. The average degree of a node or a set of nodes is noted as ($\bar{d}$). | It reflects the number of competitions a vehicle has in the network. The average degree indicates the average number of co-consideration relations for a vehicle model or brand. As shown in the following example, there are two Ford vehicles – one has degree $d_1 = 2$, and the other has $d_2 = 4$. Ford band has an average degree $\bar{d}_{ford} = \frac{6}{2}$. |
| Number of neighboring nodes ($n$) | The number of nodes directly connected to a node or a set of nodes in the same group. The average number of neighboring nodes is noted as ($\bar{n}$). | It measures the size of a vehicle co-consideration set for a vehicle or a vehicle brand, which implies the total number of competitors of a vehicle or a vehicle brand. As shown in the following example, the Ford vehicle 1 has two co-considered vehicles, $n_1 = 2$. The Ford brand has four co-considered vehicles, $n_{ford} = 4$. |
| External degree ($d'$) | The number of links that connect a node to nodes in different groups. The average external degree of a set of nodes in the same group is noted as ($\bar{d}'$). | The external degree reflects the competition between brands. In the example, the Ford vehicle 1 only connects 1 non-Ford vehicle, thus $d'_1 = 1$. Similarly, $d'_2 = 3$. As there are two Ford vehicles, the average external degree of Ford brand is therefore $\bar{d}'_{ford} = 4/2$. |
| Global clustering coefficient ($c_G$) | The proportion of closed triplets over all possible triplets (both open and closed) in a network. It is an indicator of the level of three-way competitions in a network. | $c_G$ measures the cohesion or segmentation of the vehicle market. A low $c_G$ network could have many open triplets (stars), e.g., a vehicle co-considered with many other unconnected vehicles; while a high $c_G$ network has many closed triplets (triangles), e.g., any group of three vehicles are co-considered with each other. The following example shows three closed triplets and 13 open triplets, thus $c_G$ is low. |
| Local clustering coefficient ($c$) | The likelihood that two neighbors of a node are also neighbors with each other. In other words, the percentage of all 2-stars that are closed 3-cycles. The average local clustering coefficient of a given set of nodes is noted as ($\bar{c}$). | A vehicle with a high $c$ is usually embedded in one cohesive vehicle segment where its competitors are also frequently being co-considered (structures of triangles/cycles); a vehicle with a low $c$ may have dissimilar competitors in many different vehicle segments (structure of stars), e.g., family sedans have higher $c$ than that of crossover SUVs. In the example below, Ford Chang'an Ecosport is not involved in any triangular relationships with other vehicles, so $c_1 = 0$. |
| Betweenness centrality ($b$) | The number of times a node acts as a bridge along the shortest path between two other nodes. | Vehicles with high betweenness centralities are most likely at the boundary between different market segments, e.g., a crossover SUV typically has higher betweenness than a classic sedan. In the example, to reach the other five vehicles, GM Buick Encore has to pass Ford Chang'an Ecosport five times, so $b_1 = 5$. Similarly $b_2 = 9.5$ indicates that Ford Chang'an Focus is likely at the boundary of a cluster. |

| Example of vehicle co-consideration network | Network metrics |
|---|---|
|  | $d_1 = 2, d_2 = 4$ <br> $n_1 = 2, n_{ford} = 4$ <br> $\bar{d}_{ford} = \frac{6}{2}, \bar{d}_{all} = \frac{16}{7}$ <br> $d'_1 = 1, d'_2 = 3, \bar{d}'_{ford} = \frac{4}{2}$ <br> $c_G = \frac{3}{13}$ <br> $c_1 = 0, c_2 = \frac{1}{6}$ <br> $\bar{c}_{ford} = \frac{1}{12}, \bar{c}_{all} = \frac{3}{14}$ <br> $b_1 = 5, b_2 = 9.5$ |

## 4. Case study – vehicle co-consideration predictions under fuel economy-boosting technologies in the China market

This section illustrates how the proposed approach forecasts the impact of new technologies on customers' co-consideration of vehicles and market competitions. This case study uses the China auto market as an example to analyze the impact of new technologies. In recent years, various fuel economy-boosting technologies have emerged, including new combustion strategies, lighter weighting materials, series–parallel hybrid, etc. In this paper, we first evaluate the market response to the reduction of fuel consumptions, a direct outcome of various fuel economy-boosting technologies. We then pick a specific fuel economy technology – the downsized turbo engine – that has an impact on attributes including engine power, fuel consumption, turbo, and engine size, to predict the market response of turbocharged vehicles relative to traditional gas-powered vehicles. The obtained results are useful in three aspects: (a) for understanding the underlying product attributes that determine customers' co-consideration decisions, (b) for analyzing the competitions between different vehicle models and brands, and (c) for guiding auto companies to create marketing plans and product design strategies in preparation for new technology scenarios.

### 4.1. Step 1 – Data collection

The data, from a market survey for new car owners in 2013, consists of 49,921 respondents and 389 unique vehicle models in China. For each respondent, the survey lists the purchased vehicle, the main alternative vehicle, and other vehicles considered by the individual. Due to the restrictions on survey design, respondents could only list up to three vehicles in their consideration sets (including the final purchase), even though the actual number of consideration might be higher. In addition, the information about the attributes of vehicles (e.g., engine power and fuel consumption), customer demographics (e.g., age and income), and perceived vehicle characteristics (e.g., youthful, sophisticated, and business oriented) is also collected in the survey data.

### 4.2. Step 2 – Co-consideration network construction

Using the survey data, a product co-consideration network is constructed based on the vehicle consideration information. The existence of a link between two nodes (vehicles) in the network is determined by the *lift* metric, which normalizes the co-occurrence frequency of products by the mere frequency of each product in the dataset. The *lift* between vehicle models $i$ and $j$ is calculated by (6).

$$lift(i, j) = \frac{Pr\{coconsider\ i\ and\ j\}}{Pr\{consider\ i\} \cdot Pr\{consider\ j\}}. \tag{6}$$

$Pr$ is the probability of (co)occurrence. The *lift* value indicates how likely two products are co-considered by customers, normalized by the product popularity in the market. The *lift* is a value between 0 and infinity. By the statistical definition, a *lift* value close to 1 indicates that the two products appear in customers' consideration sets almost as often as expected at random. A lift value greater than 1 indicates that the two products are more likely to be co-considered than
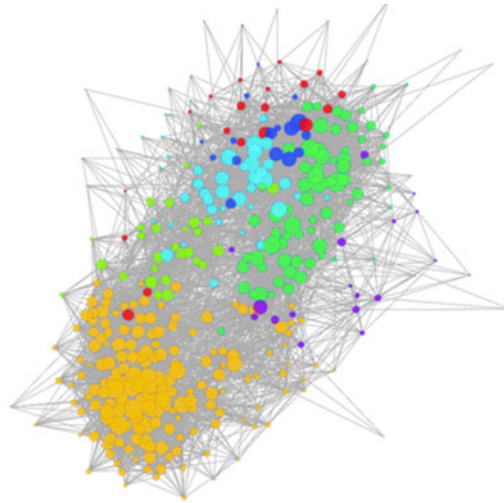
**Figure 6.** Visualization of vehicle co-consideration network. Link weights are lift values and only links with weights larger than 1 are included. Network communities are depicted using different colors.

expected by random. We have examined the lift distribution[1] and conducted sensitivity analysis on network robustness against the lift values[2]. The results revealed that the selection of lift = 1 provides a reasonable separation of the significantly, positively correlated co-considerations from others. Assuming symmetric and reciprocal of co-consideration relations, a weighted, undirected vehicle co-consideration network is produced.

Figure 6 displays the vehicle co-consideration network involving all 389 vehicles in the survey data, visualized by the Fruchterman–Reingold force-directed algorithm (Fruchterman & Reingold 1991). The colored communities are detected using the modularity maximization algorithm which optimizes the modularity score (a measure of division quality) with the greedy heuristic (Newman 2006). The algorithm is chosen because it generates non-overlapping groups, helps better visualize the market segments, and provides insights into the competition structures. The identified network communities are highly associated with the market segments. For example, the yellow community includes many domestic entry-level sedans (e.g., BYD F6 and Chery QQ), while the green community includes premium SUVs by foreign manufacturers (e.g., Jeep Grand Cherokee and Volvo XC60). However, the vehicle segment is not the only factor contributing to the formation of network communities. Product designers, for example, may ask about the shared design attributes for vehicles in the yellow community, and the common characteristics of consumers who consider these vehicles. The following step explores multiple factors together, including vehicle attributes, customer demographics, and customer perceived vehicle characteristics, to describe the emergence of vehicle communities.

[1] The lift value has a power law distribution, which means that most connections have values below 1 and around 1. Selecting a higher cutoff will aggressively omit many potential positive associations.
[2] We have tested our models using cutoff at 1, 3, 5, and 7, respectively, and no significant changes in the network characteristics or the trends of the model results are observed.
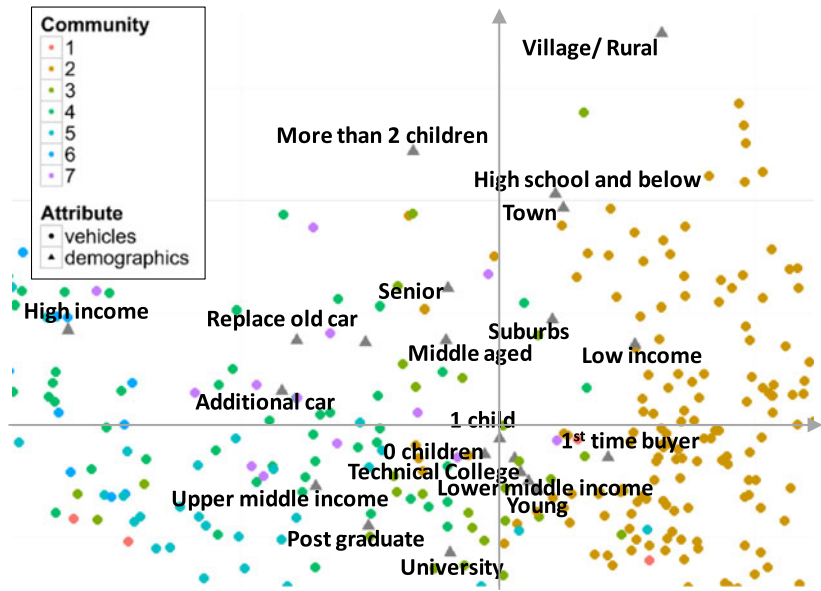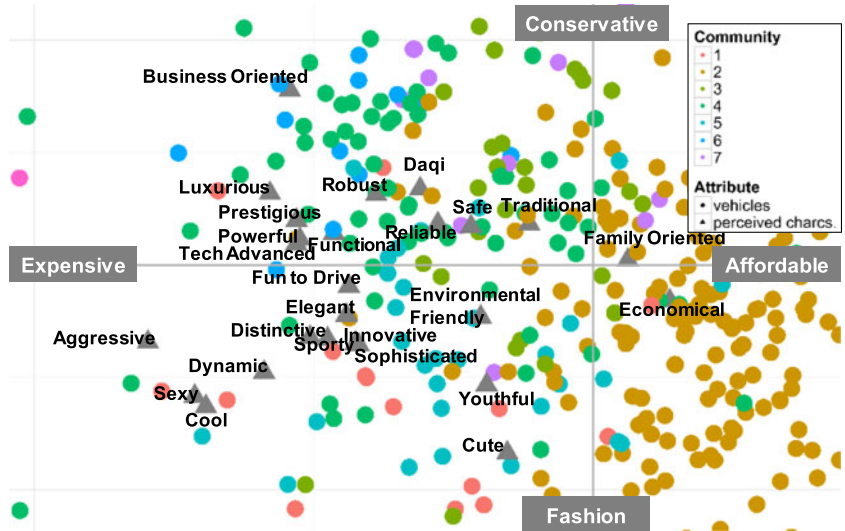
**Figure 7.** JCA plot of vehicles (in dots) and customer demographics (in triangles). The colors of the dots highlight the network communities of vehicles (representing aggregated consideration sets). As observed, the customer demographics can somewhat explain particular patterns of vehicle communities.

## 4.3. Step 3 – Joint correspondence analysis for understanding key influencing factors

In this step, we use JCA to explain the formation of network communities (aggregated consideration sets) by exploring its underlying driving factors associated with customer demographics, product attributes, and customer perceived product characteristics. The three types of attributes and their impacts on vehicle co-considerations are evaluated first separately and then together. Product attributes of vehicle models are the physical properties or specifications determined by the manufacturer or designer (e.g., brand and performance). Perceived product characteristics are collected by the subjective opinions of customers (e.g., comfortable and cool), which can be emotional and strongly influenced by the society and social media. The two types of information provide different viewpoints from designers and customers but have strong connections and interactions in between.

First, we choose the vehicle models and customer demographics as the column variables of interests. Performing JCA on the Burt matrix explains 67.3% of the total inertia using the first two dimensions. Figure 7 shows the generated joint plot. As noted, the output plot simultaneously displays all levels of the two sets of variables – the vehicle models in dots and the demographical attributes of customers in triangles. As observed from the upper left region in Figure 7, customers who have *High Income* and who acquired *Additional Car* are close to each other, and both of them are closely associated with blue dots representing luxury sedans (community #5). Also noted from the upper right of Figure 7, customers from *Village/Rural* and *Town* are also characterized by low education

16/34

**Figure 8.** JCA plot of vehicles (in dots) and customer perceived vehicle characteristics (in triangles). The colors of the dots highlight the network communities of vehicles (aggregated consideration sets). The perceived vehicle characteristics have relatively weaker power in explaining the emergence of vehicle communities.

(*High School and Below*) and associated with lower-end vehicles (community #2). A majority of customer attribute levels, however, are clustered around the principal origin, e.g., *Lower Middle Income*, *Technical College*, *Young*, *0 Children*, *1 Child*, etc., meaning that those demographic levels cannot distinguish vehicle considerations from one to another very well.

In addition to studying the relationships between variables, JCA could also help answer the question of the formation of consideration sets. More specifically, we can observe whether two vehicles within the same community form clusters and what attributes explain the formation of the vehicle communities. Figure 7 colors the vehicle points to highlight the community membership. Some communities show clear boundaries being separated from others (e.g., #2), while some communities are less clustered (e.g., #6 and #7), and even somewhat dispersed (e.g., #1, #4, and #5). The domestic, low-end sedan community (#2, in yellow) is associated with the demographics including *Village/Rural*, *Low Income*, and *High School and Below*. In contrast, the import SUV community (#4, in green) is characterized by a different set of demographics, including *High Income*, *Additional Car*, and *Replace Old Car*. Both observations are consistent with our prior understanding. For this case study, the demographical attributes can somewhat explain particular patterns of considerations, but the correlation is not strong. Such insights can stimulate the development of rigorous predictive models in the next step as well as offer opportunities for data reduction.

Second, we use vehicle models and customer perceived vehicle characteristics based on customers' subjective expressions as explanatory variables to study their impacts on vehicle co-considerations. The JCA explains 77.4% of the total inertia using the first two dimensions. Figure 8 shows the generated joint plot.

Figure 8 allows us to explore the relationships between the perceived vehicle characteristics as described by customers and the aggregated consideration sets as network communities. This effort also reveals how customers evaluate vehicles differently based on the subjective feelings and the characteristics they care most for each vehicle community. In Figure 8, the horizontal axis represents the Expensive–Cheap dimension running from the left to the right and the vertical axis represents the Conservative–Fashion dimension moving from the top to the bottom. For example, the domestic, low-end sedan community (#2, in yellow) is associated with characteristics like *Economical* and *Family Oriented*; while the import SUV community (#4, in green) is associated with characteristics like *Business Oriented* and *Prestigious*.

The JCA plot also provides design engineers with additional insights into bundling vehicle features that are compatible for a specific market niche. For example, according to the distance of triangles, environmental-friendly cars could be designed with bundling of innovative features, sophisticated look, reliable components, and safety equipment; while family-oriented cars may be designed by combining traditional exterior designs, lower prices, and fuel-efficient technologies. JCA-based qualitative analysis enables us to gain such design insights.

Next, we use vehicle models and vehicle attributes as explanatory variables to study their impact on vehicle co-considerations. In Figure 9, the relationships among the two sets of variables are much clearer to see in an inverted double V-shaped pattern, where vehicle attributes are widely distributed in the space and vehicle models are positioned close to the vehicle attributes nearby. As noted, vehicle points in the figure are clustered by network communities, meaning that vehicle models in the same community mostly share the same set of vehicle attributes. Even though the community boundary is clear, these results using all vehicle attributes in the data seem over-fit the consideration relations. As shown, the yellow community (#2) is cut apart into three blocks, and the green community (#4) is divided into two sections. Some of the links within a community are broken by the detailed description of the product attributes. The principal inertia for the first two dimensions is low in this case – only 14.9% of the total observed variance is explained because there are so many dimensions of product attributes and the first two principal axes cannot efficiently capture the full relational patterns. This result indicates that an examination on individual product attribute is needed as what we will perform in Step 4.

Taken all the three sets of variables into account, we conduct a JCA on vehicle models, customer demographics, perceived vehicle characteristics, and vehicle attributes altogether. The resulting joint plot is shown in Figure 10 where the first two dimensions ($x$ and $y$ axes) together explain 19.0% of the total variance in the data. It is noted that Figure 9 and Figure 10 have similar graphical patterns, representing the higher importance of vehicle attributes relative to other attributes in explaining co-consideration relations. In addition, it is observed that the customer demographics and perceived vehicle characteristics concentrate around the origin, while the vehicle attributes are more dispersed in the space. This suggests that part of the previously observed relations in Figure 7 and Figure 8 can be described using vehicle attributes. One example could be that *High Income* attribute of customers and *Luxury* characteristics are closely related to the *High Price* attribute of vehicles.

**Figure 9.** JCA plot of vehicles (in dots) and vehicle attributes (in triangles). The colors of the dots highlight the network communities of vehicles (representing aggregated consideration set). Vehicle brands and origins are hidden for simplicity. The vehicle attributes have the strongest power among the three sets of variables in explaining the emergence of vehicle communities.

The JCA plot including all variables may not be the best way to explain the consideration patterns. However, compared to other traditional methods, JCA is useful in exploratory data analysis to identify systematic relationships between various types of attributes and the associations of vehicles when there is no *a priori* knowledge as to the nature of those relationships. JCA provides a useful interpretative tool to understand the relationships between various sets of connections and attributes. The revealed complex relationships would not be detected in a series of pairwise comparisons by traditional statistical approaches. For example, the analyses above clearly illustrate the distinction between the vehicle attributes and the perceived vehicle characteristics. Vehicle attributes produce a JCA space which accurately differentiates various vehicle models. However, it is difficult to explain the community differences (aggregated consideration set) using one or two latent factors, most likely due to the effect of information overloading. Regular customers cannot comprehend the complex product information and make corresponding decisions. In comparison, vehicle perceived characteristics based on customers' subjective feelings (probably influenced by branding and marketing activities) only have a few underlying dimensions, such as prices and styles discussed above. Although the perceived vehicle characteristics are weak in segmenting vehicle models, they can explain more than three-quarters of the variance of the relational data. This suggests that
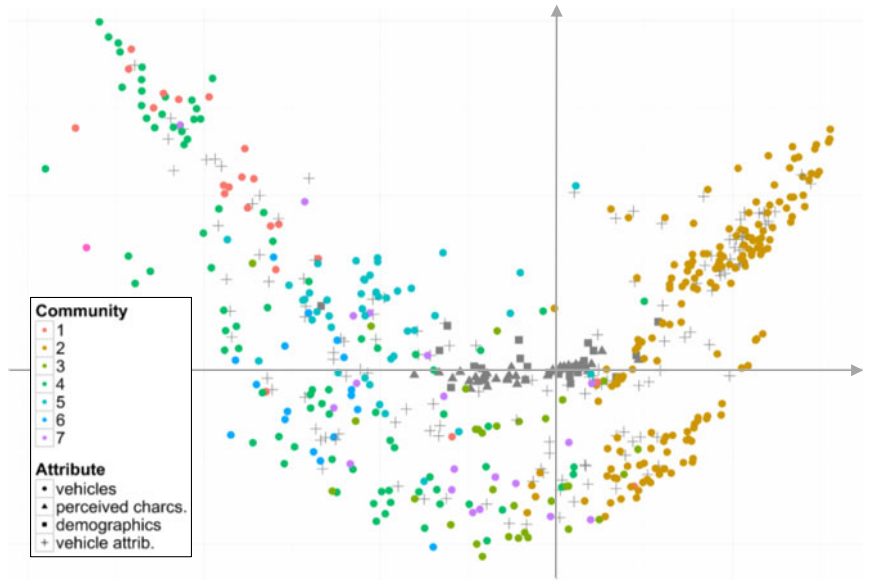
**Figure 10.** JCA plot of vehicles, customer demographics, perceived vehicle characteristics, and vehicle attributes. Different sets of variables are depicted in different shapes. The colors of the dots highlight the network communities of vehicles (representing aggregated consideration sets).

the principal dimensions generated by the perceived vehicle characteristics truly reveal how people make preference decisions.

As shown in this step, using JCA we generate visual representations that describe the communities of vehicle models, construct latent factors using correlated information, and provide a better understanding of what customer/product attributes impact the structures of vehicle communities or vehicle consideration sets. Beyond visual descriptions and explorations, in the next section, our interest is to further quantify the important effect of multiple factors on the formation of co-consideration relations of products.

## 4.4. Step 4 – MRQAP modeling

We employ MRQAP to analyze the underlying factors driving product co-consideration using a set of explanatory effect networks created by vehicle attributes and customer demographics. In the established model, the explanatory networks are built from variables of all four different types of effect networks as shown in Table 2. The two distance networks (*characteristics dist.* and *demographics dist.*) are built using the product coordinates derived from JCA in Step 3. This unique feature of the proposed method integrates the JCA with the MRQAP technique.

In data processing, a few high correlations are observed among the vehicle attributes related to price, power, and engine size. To handle this situation, we employ Dekker's double semi-partialing (DSP) method during model estimation. This method has the capability to generate relatively more robust estimates under the situation of multicollinearity (Dekker *et al.* 2007).

**Table 4.** Estimation results of MRQAP network model.

| | Attribute Type | Coeff. $\beta$ | Odds ratio $\exp(\beta)$ | Pseudo $p$-Value |
|---|---|---|---|---|
| (Intercept) | | −2.784 | 0.062 | 0.00 |
| **Vehicle Attribute Network** | | | | |
| Drivetrain match | Categorical | 0.310 | 1.364 | 0.00 |
| Gearbox match | Categorical | 0.482 | 1.619 | 0.00 |
| Fuel type match | Categorical | −0.070 | 0.933 | 0.56 |
| Brand match | Categorical | 0.625 | 1.868 | 0.00 |
| Segment match | Categorical | 1.299 | 3.666 | 0.00 |
| Vehicle origin match | Categorical | 0.284 | 1.329 | 0.00 |
| Brand origin match | Categorical | 0.619 | 1.857 | 0.00 |
| Price diff. | Numerical | −5.633 | 0.004 | 0.00 |
| Price sum | Numerical | 1.747 | 5.735 | 0.00 |
| Power diff. | Numerical | −1.791 | 0.167 | 0.00 |
| Power sum | Numerical | 0.430 | 1.537 | 0.37 |
| Fuel consumption diff. | Numerical | −3.925 | 0.020 | 0.00 |
| Fuel consumption sum | Numerical | 0.170 | 1.185 | 0.59 |
| Engine size diff. | Numerical | 0.666 | 1.946 | 0.08 |
| Engine size sum | Numerical | −0.105 | 0.901 | 0.82 |
| Turbo match | Binary | 0.350 | 1.419 | 0.00 |
| Turbo sum | Binary | 0.206 | 1.229 | 0.04 |
| **Perceived Vehicle Characteristics Network** | | | | |
| Characteristics dist. | Non-vehicle related | −0.469 | 0.626 | 0.07 |
| **Demographics Network** | | | | |
| Demographics dist. | Non-vehicle related | −0.856 | 0.425 | 0.00 |
| **Overall model fit** | | | | |
| Adjusted Pseudo- $R^2$: 0.49 | | | | |

From the results shown in Table 4, most attributes (except *fuel type match*, *power sum*, *fuel consumption sum*, *engine size diff*, *engine size sum*, and *characteristics dist.*) are statistically significant at a significance level of 0.05. This indicates that the associations formed by these vehicle attributes are important in explaining customers' co-consideration behavior. All coefficients have expected signs: the positive coefficient for a *match network* (constructed based on binary or categorical attributes, e.g., *brand match* and *turbo match*) indicates that vehicles sharing the same attribute categories are more likely to be co-considered (homophily); the negative coefficient for a *difference network* (constructed based on continuous attributes, e.g., *price diff* and *power diff*) indicates that the smaller the difference in attribute values, the more likely are two vehicles co-considered (homophily); the positive coefficient for a *sum network* (constructed based on binary or continuous attributes, e.g., *price sum* and *turbo sum*) implies that higher combined attribute values increase the probability of vehicle co-considerations (attributed-based main effects); the negative coefficient for a *distance network*

**Table 5.** Prediction accuracy of MRQAP model.

| Actual | Predicted | | |
|---|---|---|---|
| | Connected | Not connected | Accuracy |
| Connected | 1629 (33.5) | 4820 (33.5) | Sensitivity: 0.253 (0.0052) |
| Not connected | 4818 (62.2) | 64199 (62.2) | Specificity: 0.930 (0.0009) |

Notes: Evaluated by the mean of 100 simulated networks. The standard deviations are shown in parentheses.

(constructed based on non-product related attributes, e.g., *characteristics dist.* and *demographics dist.*) shows that the further the two vehicles are away from each other in the space of JCA, the less similar the non-product related attributes are shared by the two vehicles, and the less likely those two vehicles are co-considered by customers. As all the input variables are normalized, their coefficients are comparable. The magnitude explains the level of importance of an effect network, indicating how close the structure of the attribute-based effect network relates to the vehicle co-consideration network. For example, *price diff.* has the strongest effect, meaning that the difference of car price is the most influential factor in customers' co-consideration decisions. Similar to the traditional logistic regression, one can also interpret the coefficients in terms of odds ratios. For instance, the coefficient of *brand match* informs that there is an 87% increase in the odds for same-brand vehicles to be co-considered compared to two vehicles of different brands.

Different from a utility-based DCA model that only captures the main effects of product attributes, MRQAP compares different products in consideration by creating relational links through associations of attributes. This capability is especially critical to understanding product competitions because it allows answering questions related to the homophily effects, e.g., whether customers are more likely to co-consider similarly priced products.

To validate the MRQAP model's in-sample predictability, we regenerate the vehicle co-consideration network using the predicted probability of links given by (1). After 100 network simulations, we evaluate the average prediction accuracy using two measures: (a) *sensitivity*, the percentage of correctly predicted links among all actual connections; and (b) *specificity,* the percentage of missing links that were correctly predicted as such (Altman *et al.* 1994). From Table 5, we can observe the model performance with the specificity at 0.93 and the sensitivity at 0.253. The relatively low sensitivity is due to the low density of the observed co-consideration network and this is common for all link prediction applications. The actual network has 6,449 co-consideration relations out of a total of 75,466 pairs of products. This means, only 8.5% of all prediction cases have positive outcomes. This low positive rate makes the link identification to be challenging. The MRQAP method is included to illustrate the step of link prediction. More advanced methods based on similarity and shortest path (Liben-Nowell & Kleinberg 2003), hierarchical and community structures (Clauset, Moore & Newman 2008), and global network structures (Perozzi, Al-Rfou & Skiena 2014) can be used to improve the prediction performance.

### 4.5. Step 5 – Scenario formulation and network prediction

To examine the impact of technological changes on market, we make the following assumptions: (1) The market response only changes as a result of new technologies, e.g., *fuel consumption* variable in the MRQAP model, while the rest of the variables are unaffected; (2) The target population of customers remains the same as the profile distribution drawn from the survey data; (3) The new technology is only introduced for a specific set of vehicles in the market, and designs of other vehicles do not change. The following 'what if' scenarios are studied:

- *Scenario 1*: We study the general effect of fuel economy-boosting technologies by varying *fuel consumption* from 100% to 50% of its original value (at a step of 5%).

- *Scenario 2*: We study the effect of a downsized engine with a turbocharger installed. By maintaining the same power output, the turbocharged version reduces *fuel consumption* by 20% and *engine size* by 30% (Turbocharging 2015).

Note that we propose the two hypothetical scenarios for a preliminary, exploratory study. This means the scenario may not be realistic, e.g., the reduction of fuel consumption to 50% of the current capacity is difficult to achieve. Our goal is to demonstrate the potential extreme impact that the change of product attributes may bring. Besides these scenarios, we further exemplify the what-if analysis with two specific auto companies, Toyota and Ford. We are particularly interested in these two companies because of their prominent differences in the number of vehicle models available in the China market, which would result in a contrast of analysis results. For example, our data indicates that Ford has nine different vehicle models, while Toyota has 17 models. Toyota has been the leader of the fuel economy for many years, while Ford is an early adopter of the turbocharged engines across its lineup. Under the above scenarios, we investigate the impacts from two different perspectives: the full vehicle co-consideration network from a global perspective, and the networks centered around Toyota vehicles and Ford vehicles,[3] respectively, from a local perspective. The behaviors of the two local networks (centered around Toyota and Ford) are considered independently. The supporting evidence for this assumption is that there are no strong links (measured by *lift* values) between Toyota and Ford vehicles in the constructed vehicle co-consideration network.

Since the network generation process with the MRQAP model is probabilistic, before performing detailed scenario analysis, we evaluate the variation of predicted networks based on some of the network metrics proposed above (including global clustering coefficient $c_G$, average local clustering coefficient $\bar{c}$, average degree $d$, and external degree $d'$). Table 6 lists the corresponding average and standard deviation (in parentheses) of each metric computed over 100 network simulations. Both the full vehicle co-consideration network and local networks centered around a brand (Toyota and Ford, respectively) are analyzed. The small standard deviations in Table 6 imply our model is capable of predicting the vehicle co-consideration network consistently.

---

[3] We only focus on vehicles with Toyota or Ford brand. This means Lexus, for example, is not in the scope of analysis even though it belongs to Toyota company.
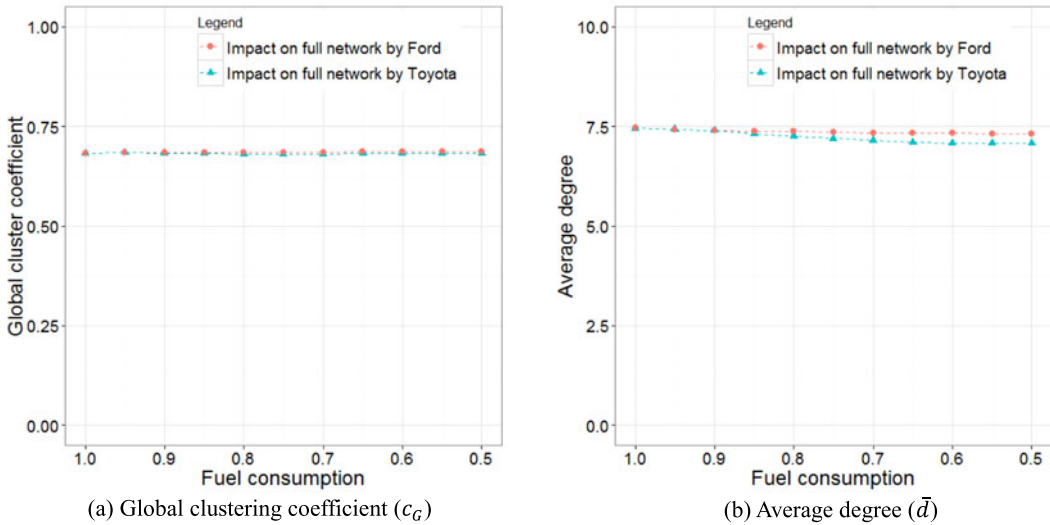
(a) Global clustering coefficient ($c_G$)

(b) Average degree ($\bar{d}$)

**Figure 11.** The impact of fuel consumption on full network.

**Table 6.** Predicted metrics averaged by 100 network simulations, the standard deviations are shown in parentheses.

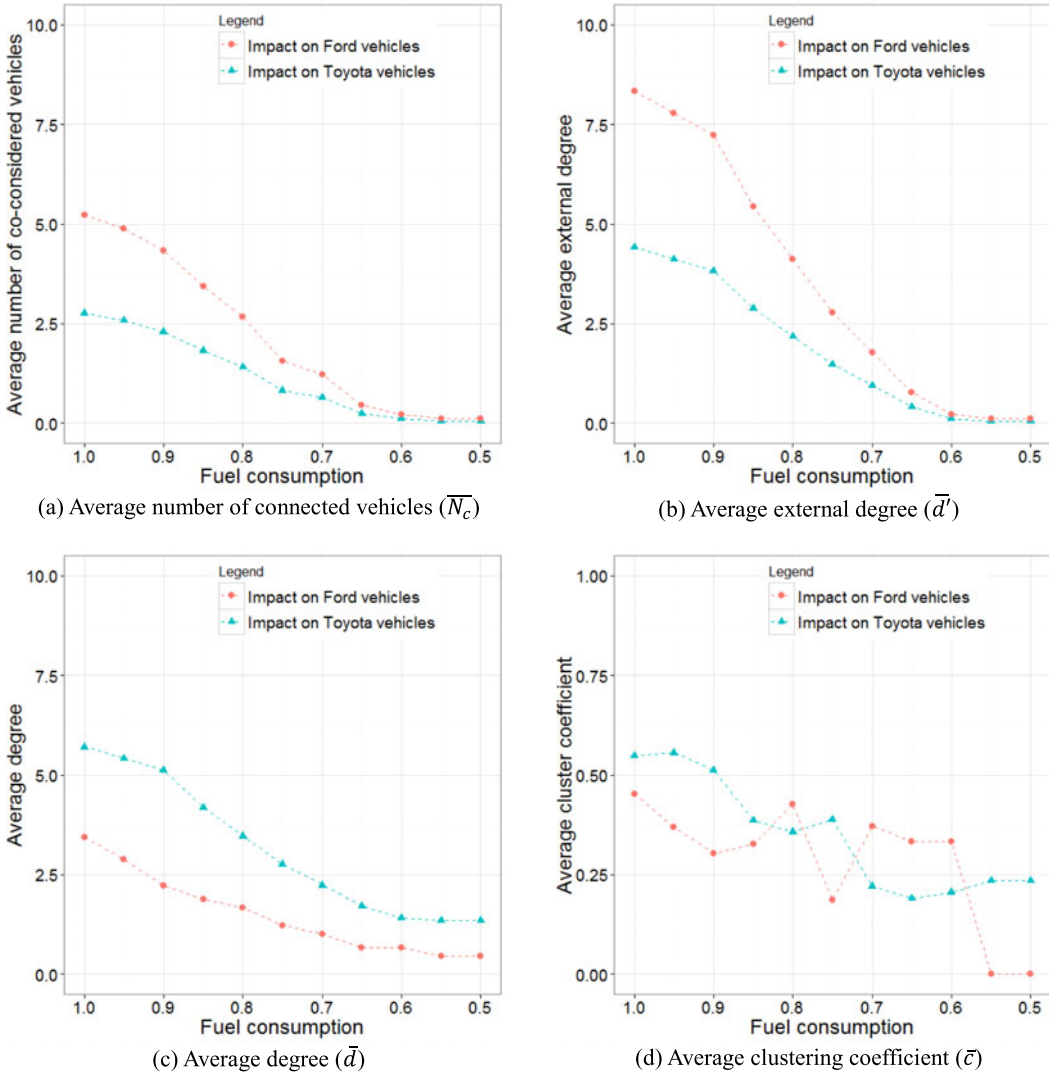| Full vehicle network | | | Toyota | | | | Ford | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $c_G$ | $\bar{c}$ | $\bar{d}$ | $N_c$ | $\bar{c}$ | $\bar{d}$ | $d'$ | $N_c$ | $\bar{c}$ | $\bar{d}$ | $d'$ |
| 0.17 | 0.18 | 33.14 | 266.32 | 0.16 | 32.06 | 499.95 | 197.3 | 0.15 | 32.57 | 277.63 |
| (0.0029) | (0.0028) | (0.41) | (7.43) | (0.0066) | (1.21) | (20.15) | (8.49) | (0.003) | (1.75) | (14.64) |

## 4.6. Step 6 – Network evaluation under technological impacts

The purpose of this step is to evaluate and interpret the change of network structures due to the technological impacts. We first apply Scenario 1 where fuel economy-boosting technologies are adopted by Toyota and Ford, respectively. In the full network (market-level) analysis, little change has been found for global clustering coefficient $c_G$ when fuel consumption decreases (Figure 11(a)). There is a slight decrease in the average degree $\bar{d}$, from 7.46 to 7.07 when Toyota reduces fuel consumption to half, and to 7.32 when Ford does the same (Figure 11(b)).

These results suggest that the application of fuel economy-boosting technologies by a single company may not affect the overall market segmentation, but may slightly reduce the competitions in the whole market. A larger impact on $\bar{d}$ (with faster decreased curve) is observed on Toyota as the brand offers more vehicle models. This implies that the market impact from one brand largely depends on the number of vehicle models that brand has.

When examining the local impact on a specific vehicle brand, we find that Ford vehicles are more likely to be co-considered with other vehicles on average (see Figure 12(a)), and especially with non-Ford vehicles (see Figure 12(b)). These results imply that in the 2013 China market, on average a Ford vehicle may have more competitors than a Toyota vehicle has.

24/34

(a) Average number of connected vehicles ($\overline{N_c}$)

(b) Average external degree ($\overline{d'}$)

(c) Average degree ($\overline{d}$)

(d) Average clustering coefficient ($\overline{c}$)

**Figure 12.** The impact of change of fuel consumption on the topology of Toyota and Ford Local Networks.

The declining trend of the two lines in Figure 12(a) shows that the average number of vehicles being co-considered decreases for both Toyota brand and Ford brand, respectively. For example, the average number of Ford's competitors reduces from above 5 to 1.25 when fuel consumption reduces to 70% of its original specification and is even 0 when the fuel consumption reaches 50%. This means, once Ford decides to adopt the fuel reduction technology, its vehicles would become more distinguishable in the market. For example, at 90% fuel reduction point, Toyota GAIG Highlander is no longer Ford Edge's rival; Honda Guangzhou Accord would not compete against Ford Changan Mondeo anymore. Please note that our analysis is performed under the assumption that the rivals' vehicle configurations are unchanged, i.e., no market game effects.

In Figure 12(b), the average external degree $\bar{d}'$ of Toyota and Ford vehicles both decline, respectively, implying vehicles from Ford or Toyota would be less likely to be co-considered with vehicles from other companies when the fuel consumption decreases. From the similar decreasing trends in Figures 12(a) and (b), one can infer that when one competing vehicle is removed (in other words, not further co-considered with Ford or Toyota vehicles), only one link is taken away. This means most of the removed links corresponded to one-on-one competition before applying the changes. It is observed from Figures 12(a) and (b) that while fuel consumption decreases, the number of vehicles co-considered for Ford decreases faster than that of Toyota vehicles. This indicates that the impacts of fuel reduction technology on Ford vehicles are more significant than the effect on Toyota vehicles in the China market.

Moreover, it is observed that the average number of links connected to Toyota vehicles or Ford vehicles, measured by average degree ($\bar{d}$), decreases as fuel economy improves (see Figure 12(c)). The declining curves indicate that the number of links taken away is bigger than the number of links added. Together with the observations in Figures 12(a) and (b), it can be inferred that the co-consideration within Toyota or Ford vehicles does not change significantly. This observation is confirmed by the data. The reason is due to the fact that the link structures in the two brand networks are predicted using the MRQAP model presented in Table 4. When the value of the *fuel consumption* variable varies, the effect networks of *fuel consumption diff.* and *fuel consumption sum* will change, accordingly. However, the attribute-based main effect represented by the *fuel consumption sum* is not significant, whereas the homophily effect represented by the *fuel consumption diff.* dominates the structural changes, as shown by its model coefficient ($-3.925$). This means when the fuel consumption for a vehicle is changed to a level significantly lower than that of the competing vehicles, the co-consideration link between them will disappear as the effect of homophily. However, the internal links within a brand largely remain unchanged, because the vehicles within the same brand are equivalently affected by the new technology and the difference of fuel consumption between two vehicles does not change significantly.

In Figure 12(d), the average clustering coefficient $\bar{c}$ of Toyota vehicles is higher than that of Ford vehicles before fuel consumption decreases. The high $\bar{c}$ implies that the competitors of Toyota vehicles are highly connected, where three-way competitions (closed triplets) appear frequently. In contrast, Ford vehicles attract more diverse competitors which are less similar with each other. Under the change of fuel consumption, $\bar{c}$ curves of Ford vehicles and Toyota vehicles decrease, respectively. The $\bar{c}$ value of Ford vehicles drops to 0 when fuel consumption reduces to 50%, as no co-considered vehicles present at that point (see Figure 12(a)), and no three-way competition exists among the Ford vehicles. The fluctuations in $\bar{c}$ curves are due to the structural variations of links both outside of competing vehicles and inside Ford vehicles. For example, a big rise in $\bar{c}$ at 80% level point can be explained by the new connection between Ecosport and New Focus. The new link makes the Ford Changan Ecosport, Ford Changan New Focus, and Classic Focus form a new three-way competition, contributing largely to the high $\bar{c}$ of Ford vehicles. The curve falls back at 75% level point, because the triangular competition formed by Ford Explorer, Toyota FAW Land Cruiser Prato,

**Table 7.** Prediction of turbo technology impacts on Toyota vehicles and Ford vehicles

| Network metrics | Toyota | | Ford | |
|---|---|---|---|---|
| | Without turbo | With turbo | Without turbo | With turbo |
| $\bar{c}_{all}$ | 0.68 | 0.68 | 0.68 | 0.67 |
| $\bar{d}_{all}$ | 7.46 | 7.14 | 7.46 | 7.39 |
| $N_c$ | 47 | 7 | 22 | 11 |
| $d'$ | 75 | 10 | 25 | 11 |
| $\bar{c}$ | 0.55 | 0.20 | 0.45 | 0.49 |
| $\bar{d}$ | 5.71 | 2 | 3.44 | 1.89 |

and Jeep Grand Cherokee is broken, resulting in Grand Cherokee being the only competitor against Explorer.

In Scenario 2, we investigate the impact of the turbo technology on vehicle co-consideration relationships. The results in Table 7 show the combined effects due to the change of turbo indicator (from 0 to 1), fuel consumption (decreased by 20%), and engine size (decreased by 30%)[4]. We find that with the turbo technology applied, the external degree $d'$ decreases for Toyota vehicles from 75 to 10 and Ford vehicles from 25 to 11, respectively, implying the declines of external competitions. As shown in Table 7, it is predicted that 40 other vehicle models (i.e., $47-7 = 40$) are no longer co-considered by customers after Toyota applies turbo technology. Similarly, the co-considered vehicles with Ford vehicles are reduced by half (from 22 to 11).

Although the number of co-considered vehicles decreases for both Toyota and Ford, respectively, the clustering coefficient $\bar{c}$ of Ford vehicles increases after Ford applies the turbo technology. This implies that the adoption of turbo by Ford may increase the connectivity (competitions) among Ford and its competitors. However, in the Toyota vehicles, $\bar{c}$ decreases from 0.55 to 0.2. Such reverse trends between Toyota and Ford indicate that same technology adopted by different vehicle brands may have different effects on the market.

## 5. Closing insights

### 5.1. Conclusions

In this paper, a network-based analysis approach is developed to facilitate the study of consideration behaviors and market competitions. Using vehicle as an example, we build a product network to establish the relationships of co-considerations from survey data. The communities emerged from such a network inform customers' co-consideration patterns in an aggregated sense. To analyze these co-consideration links, the JCA is employed to visualize vehicles associations and potential customer/product drivers. The graphical output simplifies the complex relationship structures between different sets of variables and generates an insightful description of the underlying relationships. We then develop a predictive model using multiple regressions quadratic assignment procedure (MRQAP) to provide a quantitative assessment of customers' co-consideration

[4] For vehicles already have a turbo installed, no changes in attributes are made.

relations as a function of various effect networks (match, sum, difference, and distance) created by associations of attributes. The constructed network model is capable of predicting the changes of network structures with respect to the changes of product designs and the target market. By mapping a new technology to a product attribute, the technological impacts can be forecasted on customer co-considerations as well as the market competitions. The insights and implications are crucial to identifying marketing strategies and introducing product differentiations in engineering design.

The presented network approach provides insights into the factors, e.g., price and income, which would affect customers' consideration sets. The integration of JCA and MRQAP techniques generates three important and consistent observations for the test case. First, product attributes are the most influential set of factors in customers' consideration decisions. Second, the preference heterogeneity can be partly reflected by demographics. Third, though customer perceived vehicle characteristics have the weakest explanation power to describe the co-consideration relationships, the perceived vehicle 'price' and 'style' are widely used by customers as the two major bases for decision-making.

Two scenarios of technology applications – the general fuel economy-boosting technologies and a specific turbo engine technology – are investigated. The insights drawn from this case study are summarized as follows. First, the adoption of a new technology by a single brand may not significantly change the structure of vehicle co-considerations on the whole market. Second, a new technology may lead to fewer competitors and less competition among a vehicle brand. Third, applying fuel economy-boosting technologies would reduce the competition that involves three or more vehicle models. Fourth, the same technology may bring different impacts if adopted by different brands.

## 5.2. Contributions and Limitations

The developed network model can handle complex relational data whose properties cannot be reduced to mere attributes. This capability is crucial when examining problems like co-consideration decisions where the relationship (such as similarity) between two products is possibly more important than the attributes of single products. The structure of the MRQAP model allows the evaluation of homophily effects and attribute-based main effects simultaneously. This feature is important to identifying key product attributes that drive customers' consideration decisions. This also differentiates network-based models from DCA that directly uses attributes of products and/or customers as predictors.

As a general framework, this research is a part of a larger effort to explore and address various challenges associated with complex customer–product interactions via network analysis. The major contribution of this paper is to establish the connections between consumer behavior, market competition, product design and technological development. Methodologically, this work contributes to developing a descriptive, explorative tool, and a predictive, explainable model implemented on relational, interdependent data. Nevertheless, there are a few simplifications in the model that worth attention. The co-consideration network is built using a unidimensional network that aggregates the customer preferences in co-considerations, so the information of individual heterogeneity may not be well captured. For the ease of model creation and validation, the MRQAP model and the associated scenario predictions replace

the weighted co-considerations links with dichotomized relations, i.e., either with link (1) or no link (0). This abstraction omits the strengths of co-considerations and thus the degrees of market competitions. In addition, we take consideration options equally without priorities, while in the collected survey, ordered preference on product alternatives is given by respondents. Our scenario analysis provides illustrative examples without including a complete set of explanatory variables and complex interaction effects. To better formulate the model and the scenarios, further validations with actual production data collected before and after the introduction of improved or new designs will be needed.

### 5.3. Future work

Future work needs to explore complex relational structures (bipartite and multidimensional) using advanced network modeling and prediction approaches, e.g., exponential random graph models (Wang *et al.* 2016; Lusher, Koskinen & Robins 2012) by considering not only the product attributes but also other underlying factors like preference heterogeneity and social influence. We aim to further improve the predictability while not sacrificing the interpretability of the model. Inspired by the work of Clauset *et al.* (Clauset *et al.* 2008) using domain-specific information to adjust the probabilities of edges accordingly, we have made some effort in this direction. Recently, we have been working on several different network models using product attributes, and we compared their predictive power using the same dataset. Please see our recent publication (Sha *et al.* 2017) for more details.

Inspection and prediction of weighted network connections (i.e., the degree of co-consideration) are also needed for a more comprehensive understanding of the market competitions and applications in practice. The question on the degree of co-consideration can be answered by linear regression version of MRQAP but may have different underlying mechanisms. For example, two car models with a big price difference are less likely to be co-considered. However, if the price difference is in a certain range, it may not have a significant impact on the degree of co-consideration. Moreover, a validation framework for a weighted network should also be investigated to provide performance assurance for non-binary occurrences.

Efforts will also be devoted to the robustness analysis under a variety of settings in network models, e.g., various sizes of networks, alternative measures of similarities besides *lift*, sensitivity of consideration set size on network links, and alternative data sources such as comparison engines and recommender systems, etc. Though our case study focuses on the design of fuel-efficient vehicles, the methodology can be extended to other technology-driven product design or product family design. We will also extend the network-based models to analyze purchase preferences and conduct comparative studies and benchmark testing between various network models and traditional DCA. Future efforts will also be made to introduce temporal network analysis models to understand the dynamic evolution of the network structure. One example of such technique is the temporal ERGM (Krivitsky & Handcock 2014) that can be used to study the driving factors that lead to formation or dissolution of network links from one year to another. Such dynamic network model will be capable of predicating the impact of design improvement and other real-world scenarios (e.g., launch of electrical vehicles).

## Appendix

### A.1. Algorithms of joint correspondence analysis

The MCA and JCA are extensions to simple CA approach. Because applying MCA on the Burt matrix inflates the chi-squared distances between profiles and the total inertia (Greenacre & Blasius 2006). Scale adjustments methods, such as JCA, are often applied after the unadjusted MCA solutions. To illustrate JCA, the procedures for computing unadjusted MCA solutions are presented first as follows:

(1) Compute the total of elements in **B**:

$$\mathbf{B}_{++} = \sum_{k=1}^{J}\sum_{l=1}^{J} B_{kl}. \tag{A 1}$$

(2) Divide the Burt matrix by its total:

$$\mathbf{P} = \mathbf{B}/\mathbf{B}_{++}. \tag{A 2}$$

(3) Calculate the matrix of standardized residuals **S**:

$$\mathbf{c} = \sum_{k=1}^{J} \mathbf{P}_{k-} = \mathbf{P}_{+-} = \mathbf{P}'\mathbf{1} \tag{A 3}$$

where **c** is the column mass.

$$\mathbf{S} = \mathbf{D}_{\mathbf{c}}^{-\frac{1}{2}}(\mathbf{P} - \mathbf{c}\mathbf{c}')\mathbf{D}_{\mathbf{c}}^{-\frac{1}{2}} \tag{A 4}$$

where $\mathbf{D}_{\mathbf{c}}$ is the diagonal matrix with diagonal **c**.

(4) Perform the SVD:

$$\mathbf{S} = \mathbf{V}\boldsymbol{\Phi}\mathbf{V}' \tag{A 5}$$

where $\phi_1 \geqslant \phi_2 \geqslant \cdots$.

(5) Calculate the principal inertia

$$\lambda_t = \phi_t^2 \tag{A 6}$$

and the total inertia

$$\sum_{t}\lambda_t = \sum_{t}\phi_t^2. \tag{A 7}$$

(6) Calculate the standard coordinates of columns

$$\mathbf{A} = \mathbf{D}_{\mathbf{c}}^{-1}\mathbf{V}. \tag{A 8}$$

(7) Calculate the principal coordinates of columns

$$\mathbf{G} = \mathbf{A}\mathbf{D}_{\Lambda}^{\frac{1}{2}}. \qquad \text{(A 9)}$$

where $\mathbf{D}_{\Lambda}$ is the diagonal matrix with elements $\lambda_t$ on the diagonals.

JCA is an alternative adjustment approach which uses the alternating least-squares method to correct the inflation of the total inertia. The iterative method performs repeated MCAs and adjustments until the convergence of the adjusted Burt matrix. In each iteration, the algorithm constructs an MCA approximation of the adjusted Burt matrix by changing the diagonal elements while keeping the off-diagonal elements unchanged. The detailed algorithm pseudo-code is presented as follows:

- Set $\mathbf{B_0} = \mathbf{B}$

- Repeat for $m = 1, 2, \ldots$

  * Compute MCA approximation of $\mathbf{B}^{(m-1)}$ by solving

$$\hat{\mathbf{B}}_{lk} = \mathbf{B}_{++}c_l c_k \left( 1 + \sum_{t=1}^{f} \phi_t^2 A_{lt} A_{kt} \right) \qquad \text{(A 10)}$$

where $c_l$ and $c_k$ are the column masses, $\phi_t^2$ are the principal inertia, and $A_{lt}$ and $A_{kt}$ are elements of the standard coordinate matrix $\mathbf{A}$.

- Update the main diagonal elements of Burt matrix $\mathbf{B}^{(m)}$ with the corresponding entries of $\hat{\mathbf{B}}$.

- Stop if the change of $\mathbf{B}^{(m)}$ and $\mathbf{B}^{(m-1)}$ falls below a tolerance threshold.

The JCA coordinates are computed from the converged solution $\mathbf{B}^{(\infty)}$, and the total inertia is defined as the sum of the inertias of the off-diagonal elements.

# References

**Altman, D. G.** & **Bland, J. M.** Diagnostic tests. 1: Sensitivity and specificity. *BMJ: British Medical Journal* **308** (6943), 1552.

**Benzécri, J.-P.** *L'analyse des données, vol. 2*. p. 1973. Dunod, Paris.

**Brynjolfsson, E.** 1996 The contribution of information technology to consumer welfare. *Information Systems Research* **7** (3), 281–300.

**Cantillo, V.** & **de Dios Ortúzar, J.** 2005 A semi-compensatory discrete choice model with explicit attribute thresholds of perception. *Transportation Research Part B: Methodological* **39** (7), 641–657.

**Clauset, A.**, **Newman, M. E.** & **Moore, C.** Finding community structure in very large networks. *Physical Review* E **70** (6), 066111 2004.

**Clauset, A.**, **Moore, C.** & **Newman, M. E. J.** Hierarchical structure and the prediction of missing links in networks. *Nature* **453**, 98.

**Davis, F. D. Jr.** 1986 *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. Massachusetts Institute of Technology.

**Davis, F. D.**, **Bagozzi, R. P.** & **Warshaw, P. R.** 1989 User acceptance of computer technology: a comparison of two theoretical models. *Management Science* **35** (8), 982–1003.

**Dekker, D.**, **Krackhardt, D.** & **Snijders, T. A.** 2007 Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika* **72** (4), 563–581.

**Dieckmann, A.**, **Dippold, K.** & **Dietrich, H.** 2009 Compensatory versus noncompensatory models for predicting consumer preferences. *Judgment and Decision Making* **4** (3), 200–213.

**Fiasconaro, A.** et al. Hybrid recommendation methods in complex networks. *Physical Review* E **92** (1), 2015.

**Fruchterman, T. M.** & **Reingold, E. M.** 1991 Graph drawing by force-directed placement. *Software: Practice and Experience* **21** (11), 1129–1164.

**Gaskin, S.** et al. 2007 Two-stage models: Identifying non-compensatory heuristics for the consideration set then adaptive polyhedral methods within the consideration set. In *Proceedings of the Sawtooth Software Conference*.

**Greenacre, M.** 2007 *Correspondence Analysis in Practice*. CRC Press.

**Greenacre, M.** & **Blasius, J.** 2006 *Multiple Correspondence Analysis and Related Methods*. CRC Press.

**Gilbride, T. J.** & **Allenby, G. M.** 2004 A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science* **23** (3), 391–406.

**Hauser, J. R.** 2014 Consideration-set heuristics. *Journal of Business Research* **67** (8), 1688–1699.

**Hauser, J. R.** & **Wernerfelt, B.** 1990 An Evaluation Cost Model of Consideration Sets. *Journal of Consumer Research* **16** (4), 393–408.

**Hauser, J. R.** et al. 2010 Disjunctions of conjunctions, cognitive simplicity, and consideration sets. *Journal of Marketing Research* **47** (3), 485–496.

**Heckman, J. J.** 1976 The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of Economic and Social Measurement*, vol. 5, (4), pp. 475–492. NBER.

**Hitt, L. M.** & **Brynjolfsson, E.** 1996 Productivity, business profitability, and consumer surplus: three different measures of information technology value. *MIS Quarterly* 121–142.

**Hsu, C.-L.** & **Lu, H.-P.** 2004 Why do people play on-line games? An extended TAM with social influences and flow experience. *Information and Management* **41** (7), 853–868.

**Kim, H. K.**, **Kim, J. K.** & **Chen, Q. Y.** 2012 A product network analysis for extending the market basket analysis. *Expert Systems with Applications* **39** (8), 7403–7410.

**Krackhardt, D.** 1988 Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks* **10** (4), 359–381.

**Krivitsky, P. N.** & **Handcock, M. S.** 2014 A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76** (1), 29–46.

**Lee, Y.**, **Kozar, K. A.** & **Larsen, K. R.** 2003 The technology acceptance model: Past, present, and future. *Communications of the Association for Information Systems* **12** (1), 50.

**Liben-Nowell, D.** & **Kleinberg, J.** 2003 The link prediction problem for social networks. In *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, pp. 556–559. ACM, New Orleans, LA, USA.

**Lusher, D.**, **Koskinen, J.** & **Robins, G.** 2012 *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge University Press.

**Malhotra, Y.** & **Galletta, D. F.** 1999 Extending the technology acceptance model to account for social influence: Theoretical bases and empirical validation. In *Systems Sciences, 1999. HICSS-32 Proceedings of the 32nd annual Hawaii international Conference on*, IEEE.

**McPherson, M.**, **Smith-Lovin, L.** & **Cook, J. M.** Birds of a feather: Homophily in social networks. *Annual Review of Sociology* **2001**, 415–444.

**Monge, P. R.** & **Contractor, N. S.** 2003 *Theories of Communication Networks*. Oxford University Press.

**Morrow, W. R.**, **Long, M.** & **MacDonald, E. F.** Market-system design optimization with consider-then-choose models. *Journal of Mechanical Design* **136** (3), 031003 2014.

**Mostafa, M. M.** 2015 Knowledge discovery of hidden consumer purchase behaviour: a market basket analysis. *International Journal of Data Analysis Techniques and Strategies* **7** (4), 384–405.

**Newman, M. E.** 2006 Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* **103** (23), 8577–8582.

**Perozzi, B.**, **Al-Rfou, R.** & **Skiena, S.** 2014 DeepWalk: online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 701–710. ACM, New York, USA.

**Raeder, T.** & **Chawla, N. V.** 2011 Market basket analysis with networks. *Social Network Analysis and Mining* **1** (2), 97–113.

**Resnick, P.** & **Varian, H. R.** 1997 Recommender systems. *Communications of the ACM* **40** (3), 56–58.

**Sha, Z.** et al. 2017 Modeling product co-consideration relations: A comparative study of two network models. In *International Conference on Engineering Design, ICED*, Vancouver, Canada.

**Shocker, A. D.** et al. 1991 Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing Letters* **2** (3), 181–197.

**Thatcher, M. E.** 2004 The impact of technology on product design, productivity, and profits: A duopoly model of price-quality competition. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, IEEE.

**Thatcher, M. E.** & **Oliver, J. R.** 2001 The impact of information technology on quality improvement, productivity, and profits: An analytical model of a monopolist. In *System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on*, IEEE.

**Turbocharging, R. H.** *Ray Hall Turbo Calculators*: http://www.turbofast.com.au/javacalc.html [cited 2015 11/15].

**Train, K. E.** 2009 *Discrete Choice Methods with Simulation*. Cambridge University Press.

**Venkatesh, V.** & **Bala, H.** 2008 Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences* **39** (2), 273–315.

**Venkatesh, V.** & **Davis, F. D.** 2000 A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science* **46** (2), 186–204.

**Venkatesh, V.** & **Morris, M. G.** 2000 Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly* 115–139.

**Wang, M.** et al. Modeling customer preferences using multidimensional network analysis in engineering design. *Design Science* **2**, 2016.

**Wassenaar, H. J.** & **Chen, W.** 2003 An approach to decision-based design with discrete choice analysis for demand modeling. *Journal of Mechanical Design* **125** (3), 490–497.

**Wang, M.** & **Chen, W.** 2015 A data-driven network analysis approach to predicting customer choice sets for choice modeling in engineering design. *Journal of Mechanical Design* **137** (7), 071410.

**Wang, M.** et al. 2015 Analyzing and predicting heterogeneous customer preferences in china's auto market using choice modeling and network analysis. *SAE International Journal of Materials and Manufacturing* **8**, (2015-01-0468).

**Yu, F.** et al. 2016 Network-based recommendation algorithms: A review. *Physica a-Statistical Mechanics and Its Applications* **452**, 192–208.

**Zhou, T.** et al. Bipartite network projection and personal recommendation. *Physical Review* E **76** (4), 2007.