A NETWORK APPROACH FOR UNDERSTANDING AND ANALYZING PRODUCT CO-CONSIDERATION RELATIONS IN ENGINEERING DESIGN

Wang, M., Huang, Y., Contractor, N., Fu. Y., and Chen, W.

1. Introduction

Understanding customer needs and preferences is the key for designers to offer customer-desired products (Chen, Hoyle, & Wassenaar, 2013). A recurring issue in the task of consumer preference modeling is the role of consideration decisions. If a product is ruled out by customers in consideration, the product will never be purchased. How can a designer work to perfect a product to get it into customers' considerations is an important but difficult question in product design. Often times, the consideration decision is "satisfactory" but not the "best" given the limited human (customer) information processing abilities. The resulting subset of products given by the consideration decision is referred to as the consideration set. Studies across a variety of domains, such as marketing and product design, have shown that customers' consideration sets are often small, typically in the range of 2-6 options (J. R. Hauser & Wernerfelt, 1990). To establish competitive design strategies, vehicle designers are interested in understanding how customers form their consideration sets over a large number of 300+ options in a market. Furthermore, for optimal design, quantitative models are needed to forecast changes in consideration decision as a function of changes in design attributes and features.

Existing methods such as Discrete Choice Analysis (DCA) predict customers' final choice assuming the consideration set for each consumer is given (M. E. Ben-Akiva & Lerman, 1985; Chen et al., 2013; McFadden, 1978; Train, 2009). Though DCA combines engineering models with market trends, it typically takes an additive utility form that assumes customers make *continuous tradeoffs* over different product attributes or features (the so-called compensatory decisions). However, in reality customers may not be able to evaluate or make trade-offs among a large number of attributes (Shocker, Ben-Akiva, Boccara, & Nedungadi, 1991). As expressed by the well-known funnel model (Jobber & Ellis-Chadwick, 2012), when faced with a plethora of similar products (e.g., vehicles), collecting and comparing all existing products is too labor intensive and unrealistic. Instead, customers often apply simple heuristics to explore and evaluate product options (the so-called non-compensatory decisions) (Gaskin, Evgeniou, Bailiff, & Hauser, 2007; John R Hauser, 2014). Given complex product attributes, different customer demographics, and perceived characteristics of products, as well as the correlations among them, characterizing how customers make consideration decisions and how they are influenced by product and customer attributes has been a big challenge for product designers. Our goal in this research is to develop analytical techniques to understand the connections between the formation of consideration sets and the underlying driving factors associated with product and customer attributes.

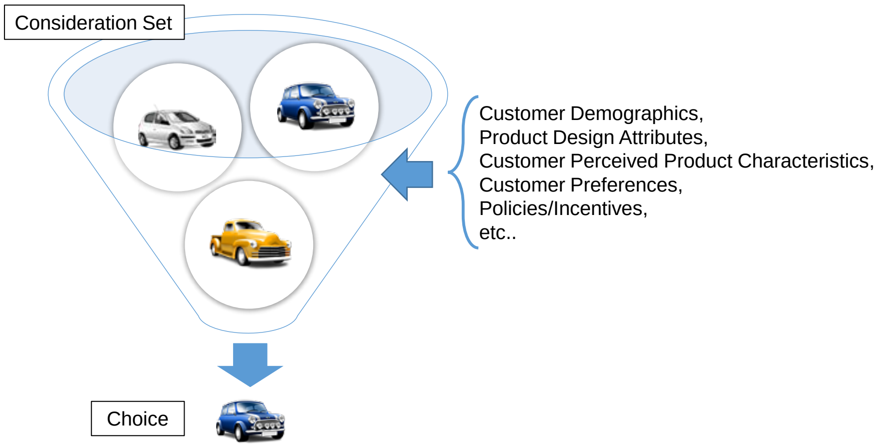

Figure 1: Illustrative representation of consideration process. Customer consideration decisions can be influenced by many factors, such as demographics, product designs, customer perceived product characteristics, customer preferences, policies/incentives, and etc.

Some pioneer works in marketing and design communities have created consideration decision models (M. Ben-Akiva & Boccara, 1995; Cantillo & de Dios Ortúzar, 2005; Dieckmann, Dippold, & Dietrich, 2009; Gaskin et al., 2007; Gilbride & Allenby, 2004; John R Hauser, Toubia, Evgeniou, Befurt, & Dzyabura, 2010; Morrow, Long, & MacDonald, 2014). Existing works in engineering design explore the suitability of various forms of non-compensatory and compensatory models and their impacts on design decision using synthetic data generated by pre-defined adjunctive rules (Morrow et al., 2014). However, these existing methods do not directly address the questions of (Q1) what products tend to be in the same consideration set; (Q2) what customer and product attributes explain the formation of the product associations; and (Q3) how the similarity (or dissimilarity) of product attributes and customer attributes impact product associations?

In this paper, we apply a novel data-driven approach based on network analysis to answer the three questions above in support of design decision making, using the vehicle consideration and design as a case study. To study what products are more likely to be considered together (Q1), we build a product association network using consumer consideration data collected from survey. To analyze how customers form consideration decisions, the Joint Correspondence Analysis (JCA) is applied to explain the formation of the vehicle associations in consideration sets (Q2). Contrast to existing preference modeling approaches such as DCA, we introduce JCA as a multivariate approach for graphical data representation, which offers a clear visual understanding of the connections between the formation of the product consideration sets and the underlying driving factors associated product and customer attributes. Finally, we employ Multiple Regression Quadratic Assignment Procedures (MRQAP) as the network version of regressions to evaluate the impact of product attributes/features and customer demographics on product co-consideration, while controlling the high interdependencies among products in the same consideration set (Q3).

This work contributes to the limited research on using network analysis to model the complex customer decision-making behaviors. Recent customer preference models in engineering design have focused on the explanation and prediction of choices between two or more discrete product alternatives, e.g., buying or not buying Toyota Prius. In contrast, in this work we explicitly model the relationships between product alternatives in a consideration set and discover the underlying driving factors to form such product co-consideration relations. All three different techniques, product association network, JCA, and MRQAP, take the “relationship” of the data among a set of products as the primary unit of analysis rather than individual observations.

The remaining part of the paper is organized as follows. Section 2 provides literature reviews and background knowledge in network analysis, JCA, and MRQAP. Section 3 introduces the dataset used for the vehicle study. Section 4 presents the method of product association network that models the connections between products in consideration sets. Section 5 details the use of Joint Correspondence Analysis (JCA) to visually describe the key drivers of product associations in consideration sets. Section 6 explains the use of Multiple Regression Quadratic Assignment Procedure (MRQAP) to predict the product co-consideration relationship as a function of similarity and difference of various product attributes and customer demographics. Finally, Section 7 offers conclusions and suggestions for future work.

2. Background and Related Work

2.1. Network analysis

Network analysis is an approach to investigate complex systems through the use of network and graph theories, using nodes as individual entities and links/ties as relationships between the entities (Wasserman & Faust, 1994). Even though the network approach has drawn increasing attention in both marketing and product design literature, the motivations in the two fields are quite different. In the product design literature, the primary use of a network was to characterize the complex architecture of a product as shared technical connections and interfaces (M. Sosa, Mihm, & Browning, 2011; M. E. Sosa, Eppinger, & Rowles, 2007). These efforts can help designers identify critical parts and product modularity in design and manufacturing. In marketing research, most applications were based on the idea of mental associative networks which connect isolated items in the form of stored knowledge (Anderson & Bower, 1973). Examples include the memory associative network of brands and the concepts, the brand associative networks (Henderson, Iacobucci, & Calder, 1998), and customers' top-of-mind associative network of products (Netzer, Feldman, Goldenberg, & Fresko, 2012). Our research objective is to bridge the gap between the marketing research and the engineering design by integrating the two types of network analysis together.

As an effort to integrate engineering design and marketing analysis using networks, a multidimensional customer-product network (MCPN) framework has been conceptually developed in our earlier research (Wang, Chen, Fu, & Yang, 2015; Wang, Chen, Huang, Contractor, & Fu, 2015) to analyze multiple types of relations involved in customer's decisions process, including the consideration and purchase decisions, social network relations, and product associations (Fig. 2). This paper focuses on the product association network which is the bottom layer of the MCPN structure. The product association links in a product network can be identified based on customer's co-consideration decisions, by connecting two products if they are frequently co-considered by customers. The product link strength reflects a mental relatedness between the two products. As a result, two products that are closely connected in the network are more likely to be compared by customers in the same consideration set.

In our previous work, various network descriptive metrics are proposed to assess the position of nodes and interpret the structural properties (Wang, Chen, Huang, et al., 2015). *Centrality* and *community* are the two primary metrics to analyze the co-consideration networks with undirected links. Centrality indicates a product's connectivity to other products and its position in the network. Measurement of centrality can be based on various properties of a node, e.g., the number of direct connections to all other nodes (degree), the minimum distance to all other nodes (closeness), and the maximum occurrence on the path of two other nodes (betweenness). Community analysis identifies products with close connections as aggregated consideration sets. Built upon our previous work that conceptually explores the meanings of network metrics (Wang, Chen, Huang, et al., 2015), this research emphasizes on searching for the physical interpretation of the network communities, and investigating their implications in product design.

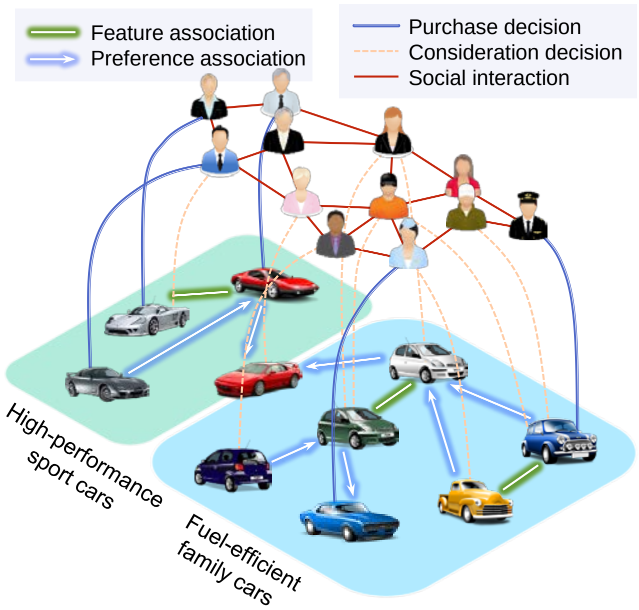

Figure 2: Multidimensional Customer-Product Network (MCPN) (Wang, Chen, Huang, et al., 2015)

2.2. Joint Correspondence Analysis

Correspondence Analysis (CA) (Benzécri, 1973; Greenacre, 2007) is a multivariate statistical technique that can be viewed as an extension of principal component analysis (PCA) applied to categorical data. The multivariate nature of CA can help in detecting structural relationships among the variable categories (i.e. levels of customers and product attributes) and objects (i.e., vehicle models, individual customers). The CA method maximizes the interrelationship between the rows and columns of a multi-way data table (data matrix) for the purpose of dimensional reduction. The data matrix often includes different sets of variables, such as vehicle attributes, customer demographics, and other derived metrics. Similar to PCA, CA creates orthogonal components and factor scores on the levels of categorical variables in the data table. The output from CA can be presented numerically or in a joint map summarizing the observed raw data and the relationships among different variables. In the generated joint map, the numerical information in the data table is transformed into a graphical display, in which each column (i.e. variable or unit of analysis) is depicted as a point in a low-dimensional space. CA is useful to map both product/customer attributes and individual products, which allows the construction of complex visual maps whose structuring can be easily interpreted. In this work, the visual map generated by the CA is used as a descriptive analysis tool which embeds the community structures of the product association network as well as the relations among customer and product attributes of interests.

The traditional Correspondence Analysis (CA) focuses on performing analysis between two sets of variables. The extension of CA to multiple categorical variables is called Multiple Correspondence Analysis (MCA) (Greenacre & Blasius, 2006). However, MCA analysis will generate different principal inertias from the CA, although the standard coordinates of the categorical variables are identical. To remedy the discrepancies in the solutions of CA and MCA, Joint Correspondence Analysis (JCA) is developed as a better generalization of CA to correct the inflation of the principal inertias (Greenacre, 2007). The implementation of JCA uses the alternating least-square method which formulates an iterative algorithm to find the optimal adjustments until convergence is achieved.

In this study, we apply CA and in particular JCA to analyze the connections between the multiple product and customer attributes and the formed network community structures. We focus on explaining the formation of network communities (aggregated consideration sets) by relating the formed network communities to the underlying driving factors associated with product attributes, customer demographics, and customer perceived product characteristics. Product attributes are the physical properties or specifications associated with the product models that are determined by the manufacturer or the designer, e.g., brand, performance, and size. Perceived product characteristics are collected by the subjective opinions of customers, which can be emotional and strongly influenced by the society and social media, e.g., comfortable, youthful, and traditional. The two types of information provide different viewpoints from designers and customers but have strong connections and interactions in between. The proposed approach generates JCA graphical outputs, constructs latent common factors using correlated information, and provide a better understanding of what customer/product attributes impact the formation of a community of vehicles or vehicle consideration sets; in other words, what vehicle attributes/characteristics customers consider when forming a consideration set. The visual plot of JCA can also be used as a preliminary data exploration tool, which eliminates irrelevant or redundant attributes before constructing more advanced data-driven network models (e.g., MRQAP) for prediction.

2.3. Multiple Regression Quadratic Assignment Procedure

Quadratic Assignment Procedure is a non-parametric bootstrapping approach, designed to test correlations and multiplexity between different networks on the same set of nodes (Krackhardt, 1988). The original form of QAP is later extended to a version of network regression, named multiple regressions quadratic assignment procedure (MRQAP). MRQAP performs an ordinary least squares regression that includes multiple network matrices to predict an outcome network (Krackhardt, 1988). The regression coefficients are calculated in standard ways, but the significance test is performed by a QAP-like permutation procedure. This is because, in a network matrix, observations in the same row or column is typically positively correlated, and this type of autocorrelation will make the standard errors and the p-values problematic. To handle non-independent observations, MRQAP estimates the standard errors using permutation of the dataset, and generates the pseudo-p-value from an empirical distribution. Recent work by Dekker, Krackhardt, and Snijders (Dekker, Krackhardt, & Snijders, 2007) evaluates the sensitivity of various types of MRQAP permutations. Even though it is found that all MRQAP tests degrade under conditions of extreme skewness and high spuriousness, using a Double Semi-Partialing (DSP) or Freedman-Lane Semi-Partialing (FLSP) with a t-statistic is the safest recommendation for operating MRQAP on general network data.

In this work, we employ MRQAP to analyze the underlying factors driving product co-consideration relations using a set of explanatory networks created by the attributes of heterogeneous product and consumer data. This approach decomposes the complexity of co-consideration relationships into a function of basic similarity (or difference) networks. Each similarity (or difference) network corresponds to one or more product attributes, customer demographics, and customer perceived product characteristics. MRQAP quantifies the contribution of each similarity (or difference) network in customer co-consideration relations, which helps identify the important factors considered by customers in making consideration decisions. The resulting fits and coefficients are calculated using the FLSP with a t-test, which provides robust results for various types of data.

3. Data Set

To demonstrate the proposed methodologies, we use the data from New Car Buyers Survey (NCBS) 2013, collected by an independent market research company (IPSOS) in mainland China. The survey was originally designed to collect information on customers' opinions on a number of different issues about their newly purchased vehicles. To ensure respondents are all new car buyers while also have adequate driving experience with their cars, all respondents are required to own the car no less than 6 weeks and no more than 32 weeks at the time of the survey. We choose this dataset as it covers a diverse set of factors, including the variables describing the (1) customer demographics (e.g., age, income, and education level) and the (2) vehicle attributes (e.g., body type, engine power, and fuel consumption), and the (3) customer perceived vehicle characteristics (i.e. subjective feelings of customers about the purchased cars). The perceived vehicle characteristics is collected by showing the respondents a list of expressions (e.g., family oriented, youthful, sophisticated, etc.), and ask respondents to select any number of items they feel applicable to their new cars. Our interest in this paper is to use the above three sets of variables to explain how consumers make vehicle consideration decisions.

With this aim, we focus on the data in which respondents identified themselves as the major decision makers in the vehicle purchase process as well as the main driver in the subsequent vehicle use. The resulting dataset contained 389 unique car models considered and bought by 44, 921 unique customers. Regarding the information on consideration sets, respondents were asked to list sequentially the car they purchased, the main alternative car they considered, and other cars they considered before making the purchase decision. In our data sample, 27.7% customers stated that they did not consider any other alternative vehicles when shopping for a new vehicle; 47.6% customers indicated that only one alternative was considered besides their chosen vehicle; and the rest of them provided two other alternatives. Due to the restriction from survey design, no respondent could list more than two other alternative vehicles in his/her consideration set even though the actual number of considered vehicles might be higher.

4. Product Association Network

The basis of constructing a product association network involves the evaluation of connections between products. In this study, we analyze the associations of products in customer's consideration set to identify patterns of co-consideration decisions. For example, given that many customers consider "Ford Edge", "Ford Kuga" and "Honda CR-V" together, we may extract the three car models and establish links between any pair of them. One limitation of using the co-occurrence frequency as a measure of association is that the association strength might be biased by how frequently a vehicle is considered (i.e. the popularity in the market). The association strength for a frequently considered vehicle will be stronger than that of a vehicle considered less frequently. To resolve this issue, we adopt the *lift* metric from the information retrieval literature (Bayardo Jr & Agrawal, 1999) to normalize the co-occurrence frequency of vehicle models by the mere frequency of each model in the dataset. The lift between vehicle model i and vehicle model j is calculated as

$$lift(i,j) = \frac{\Pr\{\text{co-consider } i \text{ and } j\}}{\Pr\{\text{consider } i\} \cdot \Pr\{\text{consider } j\}} \tag{1}$$

where Pr is the probability of a vehicle model in a consideration set, calculated based on the NCBS data. In general, the lift value indicates how likely two products are co-considered by a customer. Statistically, the lift value 1 indicates that two vehicles are completely independent; a lift value greater than 1 indicates the two vehicles occur together more frequently than one would expect by chance, and vice versa. In the example shown in Fig. 3, the lift between Edge and CR-V (1.2) is smaller than the lift between Kuga and CR-V (2.4), implying that after normalization, the association between Edge and CR-V is weaker than that between Kuga and CR-V, and CR-V is more likely to appear with Kuga than Edge in the same consideration set.

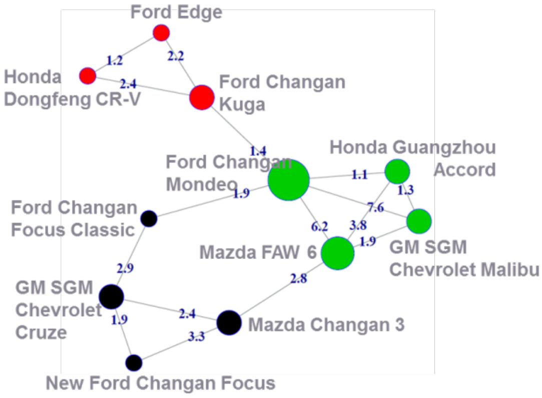

Figure 3: Illustrative network of vehicle associations

Eqn. (1) suggests that the relationship between vehicle i and vehicle j is symmetric and reciprocal. As such, we can construct a 389* 389 symmetric matrix of lifts between each pair of vehicles in the data (there is in total 389 vehicles in the dataset). The matrix is then transformed to a vehicle association network with non-directional valued links evaluated by the lift values and connected by the 389 vehicle models. Fig. 4 displays a visualization of the vehicle association network using the Fruchterman-Reingold force-directed algorithm, where nodes layout are determined by the forces pulling connected vehicles together and pushing disconnected vehicles apart. The size of network nodes represents the degree centrality of nodes (range of connections). As detailed in our previous work (Wang, Chen, Huang, et al., 2015), we applied network centrality measures to study the connection between the most connected vehicles (with a wide range of consideration against other vehicles) and the sales volume data.

To further explore vehicle associations beyond the level of vehicle pairs, we look into communities (clusters) of network nodes (vehicle models) strongly connected within one community and less densely connected with nodes in other communities and the reminder of the network. The community of network nodes suggests the aggregated consideration sets by customers because the frequently mentioned vehicles are grouped together. The typical method of community detection involves two steps: first, an objective function is defined to capture the division quality of the network as a set of nodes with better internal connectivity than external connectivity. Second, various heuristics are employed on the optimization problem to extract the interesting communities. To extract such communities of vehicle models, we employ the fast-greedy algorithm by (Clauset, Newman, & Moore, 2004), which is the most common approach in identifying network communities.
In Fig. 4, the colors of network nodes highlight the community memberships. In total, we detect seven (7) large communities where vehicles belong to the same aggregated consideration set are clustered together. An interesting observation is the high correlation between the vehicle communities and the vehicle size segments, which provides compelling evidence for the face validity of the association network approach. For example, the yellow community includes most domestic entry-level sedans (e.g., BYD F6, Chery QQ, etc.), while the green community includes premium SUVs by foreign manufacturers (e.g., Jeep Grand Cherokee, Volvo XC60, Land Rover Discovery, etc.).

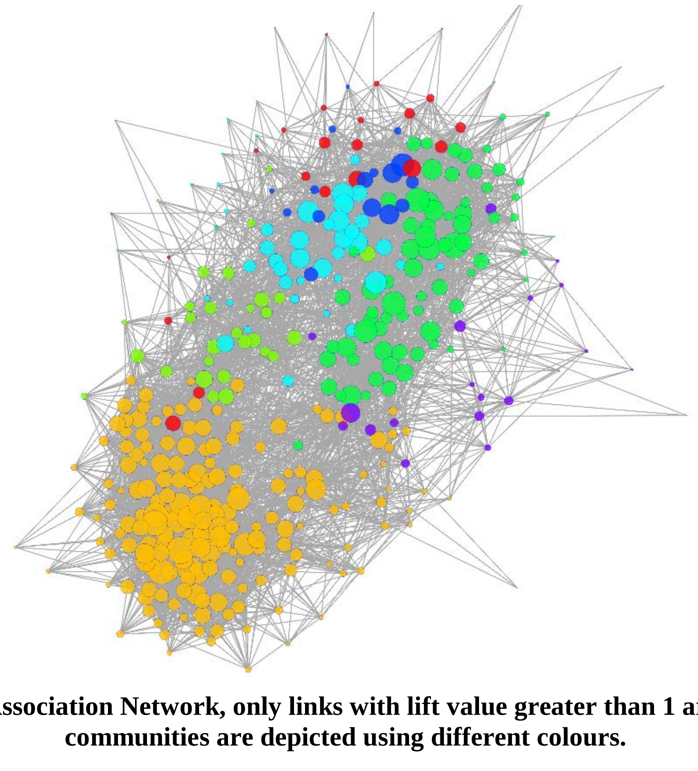

Figure 4: Vehicle Association Network, only links with lift value greater than 1 are shown. Network communities are depicted using different colours.

Because the association of vehicles in a community is not simply evoked by a single factor (e.g., vehicle size segment), but more likely is a consequence of a mixed factors. For example, as product designers, we may ask what are the common design features shared by the vehicles in the yellow community, and what kinds of consumers tend to consider vehicles in the yellow community. In the following section, we use JCA as an analytical approach to describe the emergence of vehicle communities considering multiple factors, including vehicle attributes, customer demographics, and customer perceived vehicle characteristics.

5. Joint Correspondence Analysis

5.1. Methodology

We use joint correspondence analysis (JCA) as an exploration tool to identify the underlying key product and customer attributes drivers to the network communities. Here we use the notation that is commonly seen in literature (Greenacre & Blasius, 2006) to demonstrate the JCA approach. Assume that we have $x_1,\ldots,x_q$ categorical variables (attributes) such as vehicle model and income level on N customer observations. Each x_j is associated with categorical levels $1,\ldots L_j$. Given the data, we can create a binary indicator matrix $\mathbf{Z}^{(j)}$ with $N \times L_j$ dimensions associated with each x_j. $\mathbf{Z}_{il}^{(j)} = 1$ if and only if $x_{ij} = l$. Each $\mathbf{Z}^{(j)}$ can then be concatenated to form a large indicator matrix of $N \times J$, where $J = L_1 + \ldots + L_q$ is the total number of categorical levels for all input variables x.

An example of the indicator matrix **Z** with 10 customers as row entries and two categorical variables (vehicle model x_1 and income level x_2) as column entries is shown in Table 1. The vehicle model variable x_1 shows all the available vehicles for customers to consider.

Table 1: Indicator matrix in Joint Correspondence Analysis, with customers as row entries, and vehicle models and demographical attributes as column entries.

	Vehicle model x_1			Income level x_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	1
Customer 6	0	0	1	0	1	0
Customer 7	0	1	0	1	0	0
Customer 8	1	0	0	0	1	0
Customer 9	0	0	1	0	0	1
Customer 10	0	0	1	0	1	0

Because the indicator matrix **Z** may take up large memories when the number of respondents and the number of categorical levels are large, JCA operates on the Burt matrix **B = Z'Z** which is defined as the cross-tabulation of all categorical levels. For the example above, the corresponding $J \times J$ Burt matrix is shown in Table 2.

Table 2: Illustrative Burt matrix in Joint Correspondence Analysis.

	Jetta	Camry	BMW7	Low Inc	Mid Inc	High Inc
Jetta	3	0	0	1	2	0
Camry	0	3	0	1	1	1
BMW7	0	0	4	0	2	2
Low Inc	1	1	0	2	0	0
Mid Inc	2	1	2	0	5	0
High Inc	0	1	2	0	0	3

Given the Burt matrix **B**, the coordinates of columns with respect to the principal axis can be computed by Singular Value Decomposition (SVD). The results of JCA can be then obtained by iterative updating the solution. Please see the appendix for solution details.

5.2. Results

Being a descriptive technique, CA emphasizes on the graphical representation of the results. The generated plot draws the first two dimensional of the principal coordinates of columns jointly in a Euclidean 2-D space. Using the NCBS dataset, we explore the roles of different sets of attributes in explaining the formation of network communities (i.e. aggregated consideration sets). First, we choose the vehicle models and customer demographics as the column variables of interests (as the illustrative example shown above). Performing JCA on the Burt matrix explains 67.3% of the total inertia in the first two dimensions. The generated joint map is shown in Fig.5 and Fig. 6.

As noted, the output map simultaneously display all levels of the two sets of variables - the vehicle models in dots and the demographical attributes of customers in triangles. The distances between two points in the space can be interpreted as the relative similarities in the variables examined: two vehicles are placed close to each other if they are preferred by customers with similar profiles. Two demographical attributes are grouped together if they often appear together for specific vehicle buyers; a specific vehicle and a demographical attribute are placed close to each other if customers considering the vehicle often possess the demographical attribute. It can be observed from the middle left of Fig. 6, *High Income* and *Additional Vehicle* are close to each other, and both of them are closely associated with dots representing luxury sedans (community #5). Also noted from the upper right of Fig. 6, customers from *Village/Rural* are also characterized by low education (*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 in Fig. 6, e.g., *Lower Middle Income*, *Technical/Vocational College*, *City*, *0 children*, *1 child*, etc., meaning that those demographic levels cannot distinguish vehicles from one to another very well.

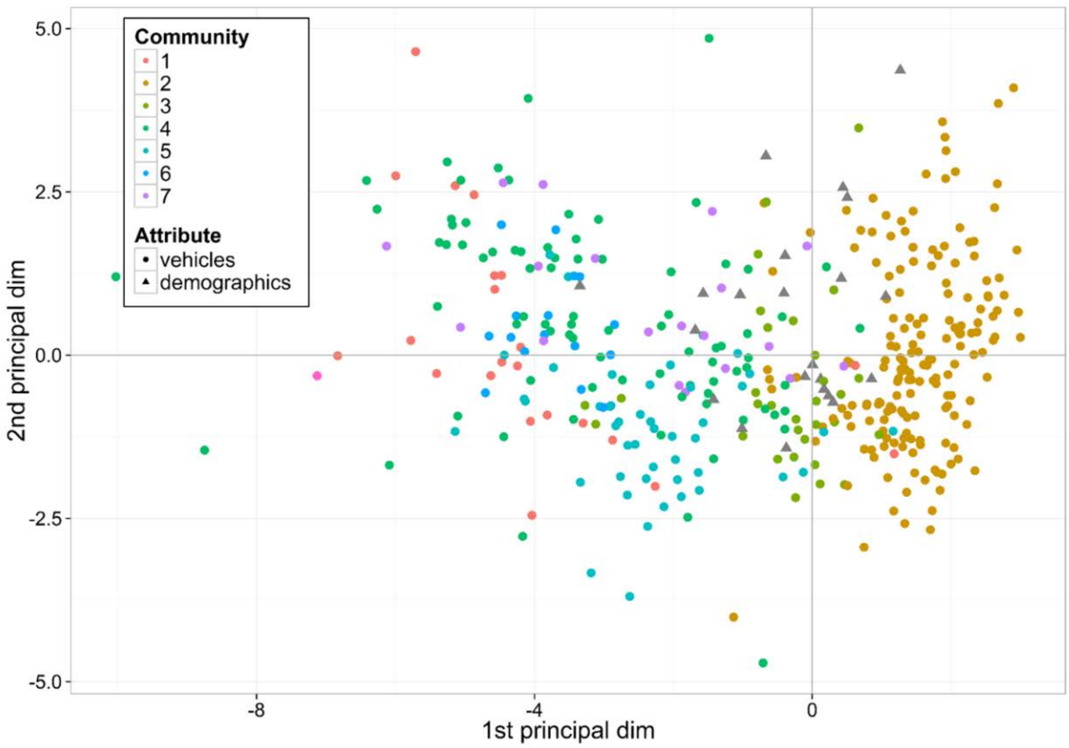

Figure 5: JCA plot of vehicles (in dots) and customer demographics (in triangles) in the first two principal dimensions. The colors of the dots highlight the network communities of vehicles (representing aggregated

consideration sets). As observed, the customer demographics can somewhat explain particular patterns of vehicle communities, but the connection is limited.

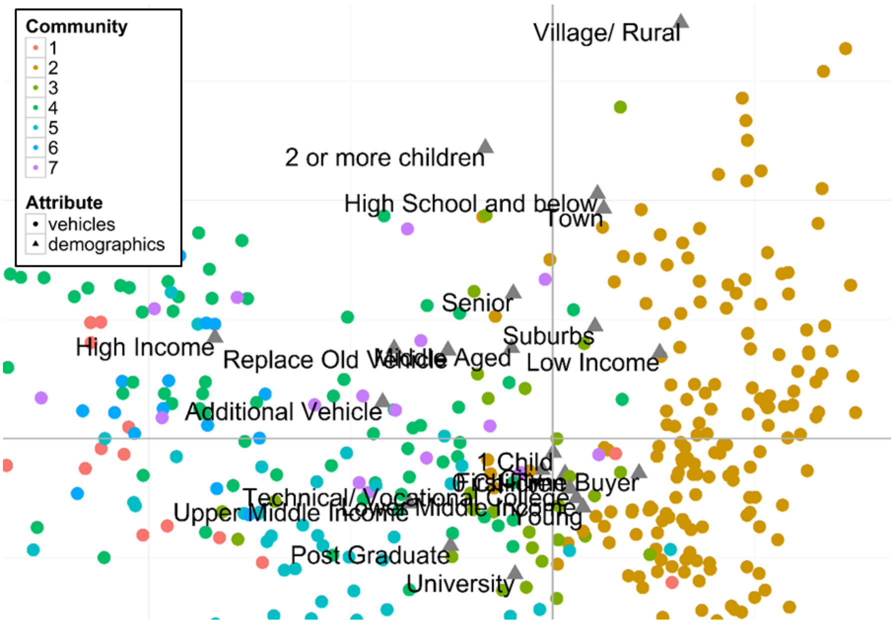

Figure 6: Enlarged JCA plot of vehicles and customer demographics.

In addition to studying the relationships between the set of variables of interest, the use of CA could also help answer the question of consideration set formation. More specifically, we can analyze whether two vehicles within the same community form clusters in the generated joint map, and what attributes explain the formation of the vehicle communities. Fig. 5 colors the vehicle points so as 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). For the domestic low-end sedan community (#2, yellow), we can see that it is featured by demographics such as “village/ rural”, “low income”, “high school and below” and etc. In contrast, the import SUV community (#4, green) is characterized by a completely different set of demographics, including “high income”, “”additional vehicle”, “replace old vehicle” and etc. Both observations are consistent to our prior understanding. For this case study, we can conclude that customer demographical attributes can somewhat explain particular patterns of considerations, but the correlation is not strong. These exploratory insights should stimulate more rigorous predictive model development 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 impact on vehicle associations. We repeat the JCA procedures and generate the results illustrated in Figs. 7-8. A numerical breakdown of the analysis shows that the first two dimensions explain 77.4% of the total observed variance in Fig. 7.

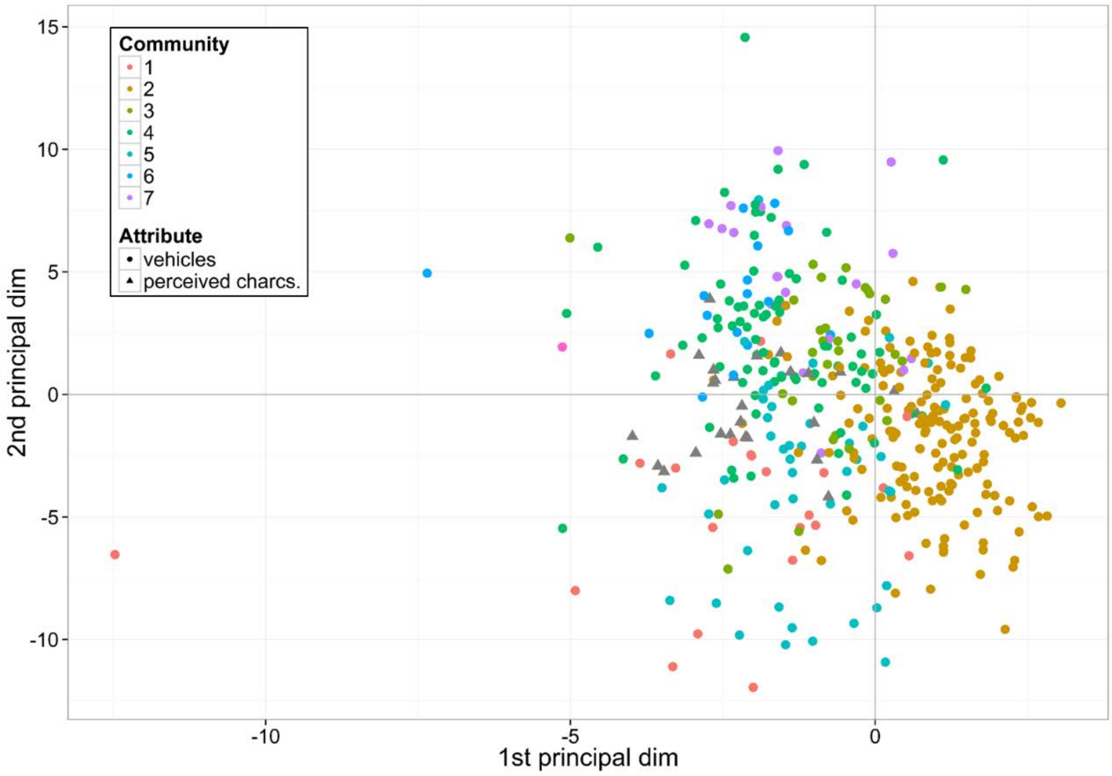

Figure 7: JCA plot of vehicles (in dots) and customer perceived vehicle characteristics (in triangles) in the first two principal dimensions. The colors of the dots highlight the network communities of vehicles (aggregated consideration sets). The perceived vehicle characteristics have relatively weaker power in explaining the emergence of vehicle communities.

Figs. 7 and 8 allow us to explore the relationships between the perceived vehicle characteristics as described by customers and the aggregated consideration sets as network communities. This effort will also reveal how customers evaluate vehicles differently based on the subjective feelings and what characteristics they care most for each vehicle community. From Fig.8, the characteristics of vehicles are relatively more crowded around the center compared to the customer attributes shown in Fig.6. Nevertheless, 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 in Fig.8. For example, the yellow community (#2) dominated by domestic low-end sedans is associated with characteristics like “economical” and “family oriented”; while the green community (#4) constituted by import SUVs is associated with characteristics like “business oriented”, “prestigious”, “luxurious”, “robust” and etc. For another example, even though the vehicles in the red community (#1) includes a very diverse car brands and classes, ranging from small size premium (Audi A5) to full size premium (Jaguar XJ), and to sports cars (Hyundai Veloster) and sports premium (Porsche 911), vehicles are relatively clustered in the lower-left area of Fig. 8 featured by semantically similar characteristics such as “cool”, “dynamic”, “sexy”, “aggressive”, “cute” and etc. To better describe the space of vehicle characteristics in the plot, one can generate a Voronoi treemap to divide the space into non-overlapping regions. The Voronoi partitioning has the property that each region contains exactly one characteristic point and every point in a given region is closer to its generating characteristic than to any other.

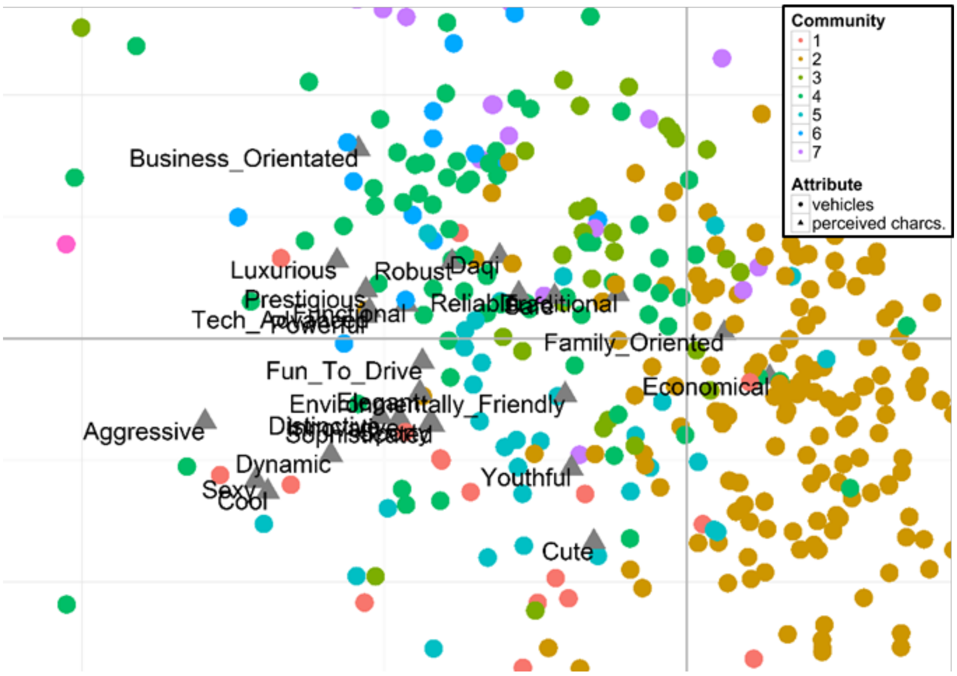

Figure 8: Enlarged JCA plot of vehicles and perceived vehicle characteristics.

Next we use vehicle models and vehicle attributes as explanatory variables to study their impact on vehicle associations. In Fig. 9, the relationships among the two sets of variables are much clearer to see: an inverted double V shaped pattern can be observed, where vehicle attributes are widely distributed in the space and vehicle models are positioned close to the vehicle attributes nearby. As noted, vehicles in the same community (color) form a cluster in the plot, meaning that vehicle models in the same community mostly share the same set of selected vehicle attributes. Even though the community boundary is clear, the model using all vehicle attributes in the data seems 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. It is also noted that the principal inertia for the first two dimensions is low in this case – only 14.9% of the total observed variance explained, because the dimension of the product attributes is so high that the JCA cannot achieve very efficient reduction.

Taken all the three sets of variables into account, finally, we conduct a JCA on vehicle models, customer demographics, perceived vehicle characteristics, and vehicle attributes altogether. The resulting joint plot is shown in Fig. 10 where the first two dimensions together explain 19.0% of the total variance in the data. It is noted that Fig. 10 is more similar to Fig. 9 than Fig. 5 and Fig. 7 in graphical patterns. Another observation is that the association between vehicle models and customer demographics, as well as the association between vehicle models and perceived vehicle characteristics, somehow disappear in Fig. 10, because most of the points for the two variables are clustered round the center. This suggests that at least part of the previously observed relations between vehicle models and customer demographics/perceived vehicle characteristics are closely related to the vehicle attributes. One example could be that *high income* customers and *luxury* vehicles are closely related to the *high price* of vehicles.

Simply looking into a 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 categories of attributes and the associations of vehicles when there are no *a priori* knowledge as to the nature of those relationships. JCA provides a useful interpretative tool that can further the understanding of relationships between the consideration sets (communities) and the socio-demographics, vehicle attributes, and perceived vehicle characteristics. The revealed complex relationships would not be detected in a series of pairwise comparisons by other multivariate statistical approaches. For example, the analyses above clearly illustrate the distinction between the vehicle attributes and the perceived vehicle characteristics. Vehicle attributes used mostly by engineers and manufacturers focus on the design perspective of the vehicles. The corresponding multi-dimensional JCA space accurately differentiates various vehicle models. However, the vehicle attributes did a poor job in explaining the variance of the relational data, most likely due to the effect of information overloading. Regular customers cannot comprehend the complex product information and make corresponding decisions. On the other hand, 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. Though the perceived vehicle characteristics have weak powers to segment vehicle models in this study, more than three quarters of the variances of the relational data are explained. This suggests the principal dimensions generated by the perceived vehicle characteristics truly reveal the rational of how people make preference decisions.

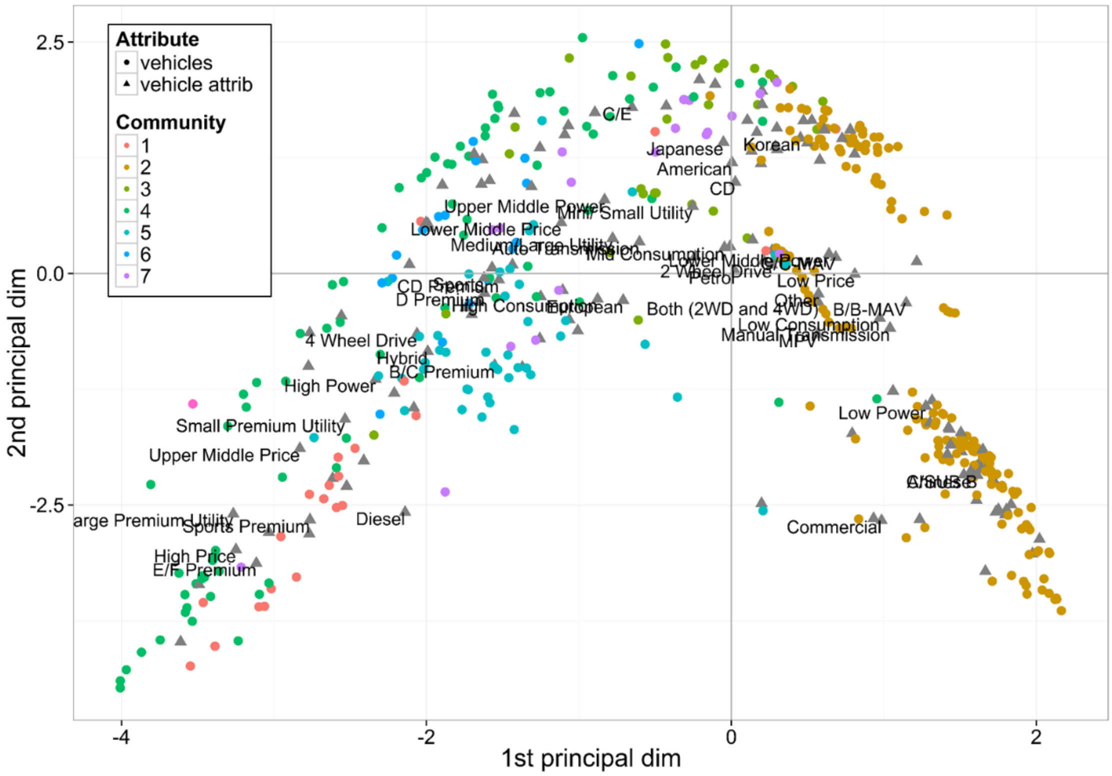

Figure 9: JCA plot of vehicles (in dots) and vehicle attributes (in triangles) in the first two principal dimensions. The colors of the dots highlight the network communities of vehicles (representing aggregated consideration set). Labels for vehicle brands and origins are hided for simplicity. The vehicle attributes have the strongest power among the three sets of variables in explaining the emergence of vehicle communities.

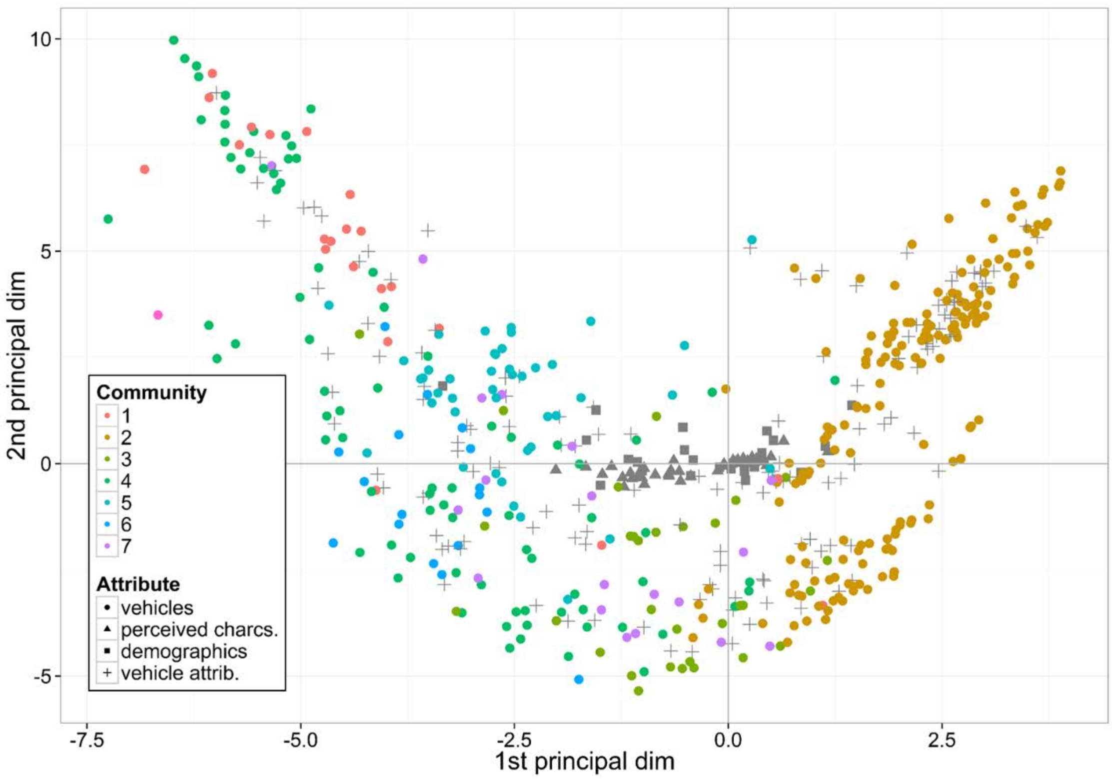

Figure 10: JCA plot of vehicles, customer demographics, perceived vehicle characteristics, and vehicle attributes. Different sets of variables are depicted in different shapes. The colors of the dots highlight the network communities of vehicles (representing aggregated consideration sets).

The network analysis and correspondence analysis presented above help us create visual representations that describe the communities of vehicles and their connections to the underlying attributes of products and customers. Beyond the visual representation, in the next section, we further quantify the importance of the multiple factors whose similarities and differences are likely to influence the formation of the consideration set. Such factors will be used to predict the change of customer consideration decisions as a result of the change of product configurations. We integrate the results of JCA to create predictive models for assessing and forecasting the association strength of vehicles in the vehicle network.

6. Multiple Regressions Quadratic Assignment Procedure

6.1. Methodology

To reveal the critical product and customer attributes that derive the structure of the vehicle co-consideration network, the technique of Multiple Regression Quadratic Assignment Procedure (MRQAP) is employed. The idea is to construct multiple similarity (or difference) networks formed based on individual product attributes and customer demographics, and then use them to predict the network structure of co-consideration relationships. The unique aspect of MRQAP is to use simple networks (created using attribute data) to predict the structure of the complex network (observed co-consideration network of vehicles). Both predictors and response are network models.

As shown in Fig. 11, the response **Y** is the matrix formed by the co-consideration lifts that determine the product co-consideration network (A-D represent product options). At the right side of the equation, the explanatory attributes are vectorized as matrix $\mathbf{X}_i$, each measures the associations among products based on the similarity (or difference) of attributes (vehicle brand, price, and the customer demographics are used as examples).

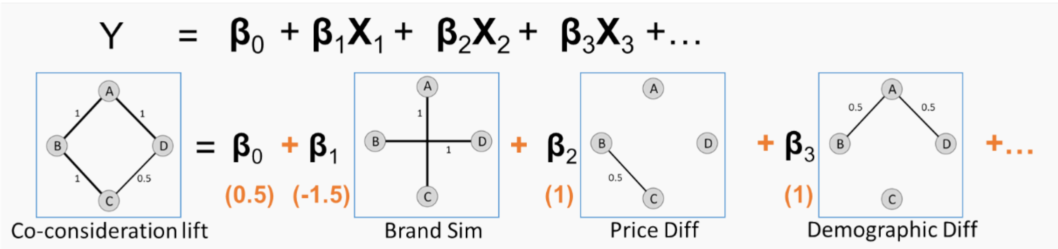

Figure 11: Illustration of MRQAP

The MRQAP analysis is employed in this study to overcome the limitation of the independence assumption in traditional Ordinary Least Square (OLS) regression, which is problematic in the case of a network (Krackhardt, 1988). Using matrix formed network data, link observations in the same row or column will be correlated. For example, each vehicle model appears 300+ times in a row or column. Although an OLS regression can provide valid point estimates, the null hypotheses tests and the associated standard errors are generally skeptical (Krackhardt, 1988). The MRQAP technique has been specifically developed to estimate the regression of network data in matrix forms. MRQAP is a two-step process where in the first step a standard OLS regression is performed to give the estimates in the usual manner. The second step involves QAP permutations of the rows and columns of a matrix followed by estimations of the OLS coefficients from the permuted matrices. The QAP permutation transforms the data matrix corresponding to a permutation of rows and columns, with rows and columns being permuted in the same way. The permutated variables are assigned new row/column numbers while all other variables are kept in the original positions. The permutation-estimation process is repeated for a large number of times, resulting in a distribution of new model coefficients that are stored away. The permutational distribution of the stored estimates is treated as the reference distribution against which the observed original estimates from the first step are compared for generating the significance values for hypotheses testing. Here we employ the FLSP approach (Dekker et al., 2007) based on the estimation and permutation of the residuals. This method has been shown to be superior under a variety of conditions of network autocorrelation, spuriousness and skewness in the data.

In creating the explanatory networks, different techniques are employed depending on how the input attribute is recorded and how the association is defined. In Fig. 11, we show examples from three vehicle attributes: price, brand and performance. In the "brand similarity" network, the association links are created based on the categorical variable *brand* to reflect if the two vehicles are of the same brand (e.g., Toyota Camry and Toyota Corolla). In the "price difference" network, the association links are built as the mean absolute deviation between the purchase prices of the two vehicles (e.g., the link value between Toyota Corolla and Ford Focus is smaller than that between Corolla and Porsche 911). In the "demographic difference" network, we use the first two principal coordinates derived from JCA in Fig. 5 to calculate the distance between the two vehicles defined by the patterns of customer demographics. The rules for creating explanatory networks in MRQAP are generalized as follows.

If the input variable is a categorical attribute, we create a similarity network:

$$X_{ij} = \begin{cases} 1 & \text{if } x_i = x_j \\ 0 & \text{if } x_i \neq x_j \end{cases} \tag{2}$$

where x_i and x_j are the categorical levels of attribute x for vehicle model i and j. If the input variable is a numeric attribute, we standardize it to a scale between 0 and 1, and then create a difference network:

$$X_{ij} = \left|x_i - x_j\right| \quad (3)$$

where $|\cdot|$ denotes the absolute value; x_i and x_j are the numeric values of the standardized attribute x for vehicle model i and j. If the input variable is represented by the coordinates derived JCA, we create a difference network:

$$X_{ij} = \left\|\mathbf{x}_i - \mathbf{x}_j\right\|_2 \quad (4)$$

where $\|\cdot\|_2$ is the L2-norm; $\mathbf{x}_i$ and $\mathbf{x}_j$ are the first two dimensions of principal coordinates **G** for vehicle models i and j from the JCA. If necessary, more dimensions can be used in the same way. Finally, for the dependent network of lifts, we make a log-transformation on its values to handle the ratio-based metric (Netzer et al., 2012):

$$Y_{ij} = \log\left(1 + lift_{ij}\right) \quad (5)$$

6.2. Results

The result of MRQAP based on the NCBS data is reported in Table 3. We compare three MRQAP models to highlight the contributions of different types of variables.

Table 3: Comparisons of the MRQAP Models. Vehicle attributes are chosen directly from NCBS data.

	Model 1		**Model 2**		**Model 3**	
	Vehicle Attributes		Vehicle Attr.+ Characteristics		Vehicle Attr. + Characteristics + Demographics	
	Coeff.	Pseudo-P-Value	Coeff.	Pseudo-P-Value	Coeff.	Pseudo-P-Value
Intercept	**0.1474**	0.000	**0.1525**	0.000	**0.1646**	0.000
Vehicle Attributes Network						
Drivetrain sim.	**0.0474**	0.000	**0.0474**	0.000	**0.0463**	0.000
Gearbox sim.	**0.0215**	0.020	**0.0217**	0.010	**0.0222**	0.010
Fuel Type sim.	-0.0223	0.140	-0.0221	0.130	-0.0218	0.140
Brand sim.	**0.2909**	0.000	**0.2912**	0.000	**0.2887**	0.000
Segment sim.	**0.4315**	0.000	**0.4309**	0.000	**0.4316**	0.000
Vehicle Origin sim.	**0.0685**	0.000	**0.0670**	0.000	**0.0653**	0.000
Brand Origin sim.	**0.1288**	0.000	**0.1298**	0.000	**0.1292**	0.000
Price diff.	**-0.1205**	0.000	**-0.1177**	0.000	**-0.1093**	0.000
Power diff.	-0.0265	0.480	-0.0256	0.530	-0.0215	0.550
Fuel Consumption diff.	**-0.2878**	0.000	**-0.2883**	0.000	**-0.2848**	0.000
Perceived Vehicle Characteristics Network						
Characteristics diff.	--	--	-0.0281	0.370	0.0022	0.980
Demographics Network						
Demographics diff.	--	--	--	--	**-0.0749**	0.000
Adjusted R-squared	0.1299		0.1300		0.1303	
The two-tailed pseudo-p-value is calculated by the proportion of times the absolute value of the original estimates are larger than the absolute value of the permuted estimates across 500 iterations. Bolded coefficients are those significant at 0.05 level.						

Model 1 formulates a MRQAP analysis that regresses the lifts of vehicle co-consideration network on the similarity (or difference) in vehicle attributes only. These attributes are mostly engineering attributes of a vehicle whose values can be later changed through design. Most attributes except *fuel type* and *power difference* have significant coefficients, indicating the relationships in these attributes among the vehicles are important in explaining the vehicle co-consideration. Consistent to the results from JCA in Sec. 5, vehicles sharing similar attributes are more likely to be co-considered, as explained by the positive relationship between a similarity network formed for product attributes and the lifts in the co-consideration network, and the negative coefficients for a difference network. By examining the values of the coefficients, we find that the vehicle attribute *segment* has the strongest effect, meaning that the structure of the *segment* network is the most closely related to the co-consideration network. *Brand* and *fuel consumption* also have moderately strong effects followed by *brand origin*, *price, vehicle origin, gearbox, and drivetrain*. While our analysis focuses on the link relationships of networks, the results cannot be obtained by a simple main effect analysis.

In Model 2, we include in the regression model the vehicle characteristics network defined by the coordinates of vehicles in JCA shown in Fig. 7. This time, we predict the co-consideration network of vehicles as a function of how similar they are in customers' perceptions in addition to the product attributes. Interestingly, the newly included *characteristics difference* has an insignificant coefficient, indicating the difference of customers' perceived characteristics has little influence on co-consideration relationships compared to the influence from real product attributes. This conclusion is also justified from the JCA results in Sec. 5.

In Model 3, we further examine how customer demographics can help explain the co-consideration between vehicles. We include the information of customers along with the perceived vehicle characteristics and real vehicle attributes in a single model. The demographics difference network is created based on the coordinates of vehicles in JCA shown in Fig. 5, where the locations of vehicles are controlled by the similarity of customer demographics. As expected, a negative and significant coefficient is identified for the *demographics difference* network. This means that vehicles preferred by customers with similar demographics (low distance in the JCA coordinates and low link value in the demographics difference network) tend to be co-considered by customers (high lift value in the co-consideration network). It is also observed that in Model 3 the nature of the effects by the vehicle attributes and characteristics does not change much from the nested models (Model 1 and Model 2), but each addition of the effect does improve the model fit. Thus, we conclude that the similarity in the vehicles' segments, brands, fuel consumptions, and prices along with the customer demographics are the dominating factors that help to explain the degree of co-consideration between vehicles.

7. Conclusions

In this study, we develop an association network approach to analyze both qualitatively and quantitatively the impact of various factors such as product attributes and consumer demographics on the formation of customers' consideration sets. Using vehicle as an example, we first build a vehicle association network from customer data of consideration sets where the link strength in network represents the tendency of co-consideration. The communities emerged from the vehicle association network informs the customer co-consideration patterns in an aggregated sense. We then employ network visualization and modeling tools to the constructed vehicle association network to explain the established co-consideration relationships among vehicles. Using the joint correspondence analysis (JCA), we are able to simultaneously visualize 389 vehicle models along with different sets of product and consumer variables of interests in a 2-dimensional graph. The graphical output remarkably simplifies the complex relationship structures between different sets of variables, and generates a simple yet exhaustive description of the underlying relationships. Such an effort would be prohibitively difficult if other traditional multivariate approaches are used. We then use the multiple regression quadratic assignment procedure (MRQAP) to predict the co-consideration relationships between vehicles as a function of similarity (or difference) networks created by product and customer attributes. The quantitative model help designers analyze the underlying important factors in product design to stimulate customers' considerations of products. Though not demonstrated in the paper, the same model can be used for predicting vehicle association network structure change with respect to the changes of product design and the target market.

The presented network approach provides insights into the factors that customers consider in forming the consideration set. The JCA and MRQAP techniques generate three consistent key observations for the test case: 1) Product attributes are the most influential set of factors in customer consideration decisions; 2) Customer demographics somewhat affect customer decision criteria and consideration preference; 3) 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 basis of decision making. These findings may have important implications to vehicle designers. For example, vehicle producers can place their bets on specific market segments and provide optimized product portfolios to get their products into customers' consideration set. They can also carry out effective strategic planning to react to the possible market volatility.

Examining how vehicles are connected in the association network and which factors impact the connections is valuable in that it can provide a better understanding of the patterns and rules underlying the vehicle associations. One could extend the application of the association network beyond vehicles to other industrial or commercial products, and broaden the analysis of consideration decision to other preference relations or physical connections. Additionally, designers can use the presented methodologies to study the market dynamics for the launch of new products or reposition products in response to future change of market.

The paper employs a unidimensional network modeling approach to study vehicle associations. In essence, the approach evaluates customers' average (aggregated) preference across the population. Advanced network modeling approaches that capture disaggregated preference behaviors of individual customers should be examined in the future. The current work has not taken into account the social influence on customers' consideration decisions while social influence may have more explanatory power when examining customer preference to new energy vehicles and luxury brands. Future work can examine the effect of "social influence" in customer decision and how designers can engineer socially influenced features into a product by introducing a multi-dimensional network structure.

Acknowledgement

The financial support from National Science Foundation (CMMI 1436658) and Ford Northwestern Alliance Project, and the data support from Ford Motor Company are greatly appreciated.

References

Anderson, J. R., & Bower, G. H. "Human associative memory": VH Winston & Sons, 1973.

Bayardo Jr, R. J., & Agrawal, R. "Mining the most interesting rules". Paper presented at the Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 1999. 145-154

Ben-Akiva, M., & Boccara, B. "Discrete choice models with latent choice sets". International Journal of Research in Marketing, 12(1), 1995, 9-24.

Ben-Akiva, M. E., & Lerman, S. R. "Discrete choice analysis: theory and application to travel demand" (Vol. 9): MIT press, 1985.

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

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

Chen, W., Hoyle, C., & Wassenaar, H. J. "Decision-based design: Integrating consumer preferences into engineering design": Springer, 2013.

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

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

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

Gaskin, S., Evgeniou, T., Bailiff, D., & Hauser, J. "Two-stage models: Identifying non-compensatory heuristics for the consideration set then adaptive polyhedral methods within the consideration set". Paper presented at the Proceedings of the Sawtooth Software Conference, 2007. 67-83

Gilbride, T. J., & Allenby, G. M. "A choice model with conjunctive, disjunctive, and compensatory screening rules". Marketing Science, 23(3), 2004, 391-406. doi:DOI 10.1287/mksc.1030.0032

Greenacre, M. "Correspondence analysis in practice": CRC press, 2007.

Greenacre, M., & Blasius, J. "Multiple correspondence analysis and related methods": CRC Press, 2006.

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

Hauser, J. R., Toubia, O., Evgeniou, T., Befurt, R., & Dzyabura, D. "Disjunctions of conjunctions, cognitive simplicity, and consideration sets". Journal of Marketing Research, 47(3), 2010, 485-496.

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

Henderson, G. R., Iacobucci, D., & Calder, B. J. "Brand diagnostics: Mapping branding effects using consumer associative networks". European Journal of Operational Research, 111(2), 1998, 306-327.

Jobber, D., & Ellis-Chadwick, F. "Principles and practice of marketing": McGraw-Hill Higher Education, 2012.

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

McFadden, D. "Modelling the choice of residential location": Institute of Transportation Studies, University of California, 1978.

Morrow, W. R., Long, M., & MacDonald, E. F. "Market-System Design Optimization With Consider-Then-Choose Models". Journal of Mechanical Design, 136(3), 2014, 031003.

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

Shocker, A. D., Ben-Akiva, M., Boccara, B., & Nedungadi, P. "Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions". Marketing letters, 2(3), 1991, 181-197.

Sosa, M., Mihm, J., & Browning, T. "Degree distribution and quality in complex engineered systems". Journal of Mechanical Design, 133(10), 2011, 101008.

Sosa, M. E., Eppinger, S. D., & Rowles, C. M. "A network approach to define modularity of components in complex products". Journal of Mechanical Design, 129(11), 2007, 1118-1129.

Train, K. E. "Discrete choice methods with simulation": Cambridge university press, 2009.

Wang, M., Chen, W., Fu, Y., & Yang, Y. "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), 2015.

Wang, M., Chen, W., Huang, Y., Contractor, N., & Fu, Y. "A Multidimensional Network Approach for Modeling Customer-Product Relations in Engineering Design". Paper presented at the ASME 2015 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, 2015.

Wasserman, S., & Faust, K. "Social network analysis: Methods and applications" (Vol. 8): Cambridge university press, 1994.

Appendix

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, we first present the procedures for computing unadjusted MCA solutions, as follows:

1. Compute the grand total of elements in $\mathbf{B}$:

$$\mathbf{B}_{++} = \sum_{k=1}^{J}\sum_{l=1}^{J} B_{kl} \tag{6}$$

2. Divide the Burt matrix by its grand total:

$$\mathbf{P} = \mathbf{B} / \mathbf{B}_{++} \tag{7}$$

3. Calculate the matrix of standardized residuals $\mathbf{S}$:

$$\mathbf{c} = \sum_{k=1}^{J} \mathbf{P}_{k\cdot} = \mathbf{P}_{+\cdot} = \mathbf{P'1} \tag{8}$$

where $\mathbf{c}$ is called column mass.

$$\mathbf{S} = \mathbf{D}_{\mathbf{c}}^{-\frac{1}{2}} \left(\mathbf{P} - \mathbf{cc'}\right) \mathbf{D}_{\mathbf{c}}^{-\frac{1}{2}} \tag{9}$$

where $\mathbf{D_c}$ is the diagonal matrix with diagonal $\mathbf{c}$.

4. Perform the SVD:

$$\mathbf{S} = \mathbf{V\Phi V'} \tag{10}$$

where $\phi_1 \geq \phi_2 \geq \ldots$

5. Calculate the principal inertia

$$\lambda_t = \phi_t^2 \tag{11}$$

and the total inertia

$$\sum_t \lambda_t = \sum_t \phi_t^2 \tag{12}$$

6. Calculate the standard coordinates of columns:

$$\mathbf{A} = \mathbf{D}_{\mathbf{c}}^{-1}\mathbf{V} \tag{13}$$

7. Calculate the principal coordinates of columns:

$$\mathbf{G} = \mathbf{A}\mathbf{D}_{\Lambda}^{\frac{1}{2}} \tag{14}$$

where $\mathbf{D}_{\Lambda}$ is the diagonal matrix with elements λ_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,…
    - Compute MCA approximation of $\mathbf{B}^{(m-1)}$ by solving

      $$\hat{\mathbf{B}}_{lk} = \mathbf{B}_{++} c_l c_k \left(1 + \sum_{t=1}^{f} \phi_t^2 A_{lt} A_{kt}\right) \tag{15}$$

      where c_l and c_k are the column masses, ϕ_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.