



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

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

Keywords: preference modeling, consideration set, network analysis, correspondence analysis, multiple regression quadratic assignment procedure

1. Introduction

Understanding customer needs and preferences is the key for designers to offer customer-desired products. A recurring issue in consumer preference modeling is the role of consideration decisions. The resulting subset of products (typically in the range of 2 to 6) given by the consideration decision is referred to as the consideration set [Hauser et al. 1990]. Existing methods such as Discrete Choice Analysis (DCA) predict customers' final choice, but not the consideration decision for each consumer [McFadden 1978], [Train 2009]. If a product is not included in the consideration set, the product will never be purchased. 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 over hundreds of products in the market, and how consideration decisions are influenced by product and customer attributes have been big challenges for product designers. Furthermore, for optimal design, quantitative models are needed to forecast changes in consideration decision as a function of changes in product attributes. Our goal in this work is to develop analytical techniques and quantitative models to understand the connections between the formation of consideration sets and the underlying driving factors associated with product and customer attributes.

Pioneer works in marketing and design communities have created models for consideration decisions [Gaskin et al. 2007], [Dieckmann et al. 2009], [Hauser et al. 2010], [Morrow et al. 2014]; however, these models 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 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 product as a case study.

To study what products are 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 visually explain the vehicle associations in consideration sets and the underlying driving factors associated with product and customer attributes (Q2). 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-considerations, while controlling the high interdependencies among products in the same consideration set (Q3). The remaining paper is organized as follows. Section

2 provides literature reviews on network analysis, JCA, and MRQAP. Section 3 introduces the data used for the vehicle study. Section 4 - 6 detail the three analytical techniques, respectively, in analyzing customer consideration decisions. 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 graph theories, using nodes as entities and links/ties as relationships between the entities [Wasserman et al. 1994]. Applications have been seen in both marketing and product design literature, although the motivations in the two fields are quite different. In product design, network analysis is used to help designers identify critical parts and product modularity in design and manufacturing [Sosa et al. 2007]. In marketing research, most applications are based on the idea of mental associative networks which connect isolated items in the form of stored knowledge [Anderson et al. 1973], e.g., brand associative networks [Henderson et al. 1998] and product associative networks [Netzer et al. 2012].

As an effort to integrate product design and marketing using networks, a multidimensional customer-product network (MCPN) framework has been conceptually developed in our earlier research [Wang et al. 2015], where multiple types of relations in customer's decisions process are involved, such as the consideration and purchase decisions, social network relations, and product associations (Figure 1). This paper focuses on the bottom-layer product network of the MCPN structure. The association links between products are identified based on customer's co-consideration decisions, by connecting two products if they are frequently co-considered by customers. Built upon our previous work that conceptually explores the meanings of network metrics [Wang et al. 2015], this research emphasizes on searching for the physical interpretation of the network links and communities, and investigating their implications in product design.

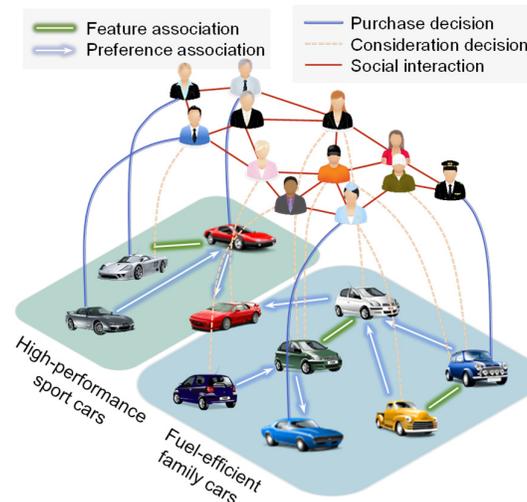


Figure 1. Multidimensional Customer-Product Network (MCPN) [Wang et al. 2015]

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. attribute levels of customers and products) and objects (i.e., products, 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. Similar to PCA, CA creates orthogonal components and factor scores on the levels of categorical variables in the data table, which allow the construction of visual plots whose structure can be easily interpreted. In this work, the visual plot

generated by the CA is used as a descriptive analysis tool which embeds the community structures of the product 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 et al. 2006]. However, MCA analysis will generate different principal inertias from the CA, although the standard coordinates of 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 products/customers attributes and the formed network community structures. We focus on explaining the formation of network communities (aggregated consideration sets) by relating the formed 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 designer, e.g., brand, performance. 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, cool. 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 plots, constructs latent factors using correlated information, and provides a better understanding of what customer/product attributes impact the structures of a vehicle community or vehicle consideration sets. The visual plot can also be used as a preliminary data exploration tool, which eliminates irrelevant or redundant attributes before constructing more advanced data-driven network models 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 this issue, MRQAP is developed to generate the standard errors and the pseudo p-value from an empirical distribution through data permutation. Recent literature [Dekker et al. 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 complexities 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 these decisions. The results the model 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. After data pre-

preprocessing, the resulting dataset contains 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. 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. 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.

4. Product association network

The basis of constructing a product association network involves the evaluation of connections between products. 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 (Figure 2, left). The lift metric [Bayardo Jr et al. 1999] is adopted here to assess the link strength, which normalizes the co-occurrence frequency of vehicle models by the mere frequency of each model in the dataset. In the example network in Figure 2(left), 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.

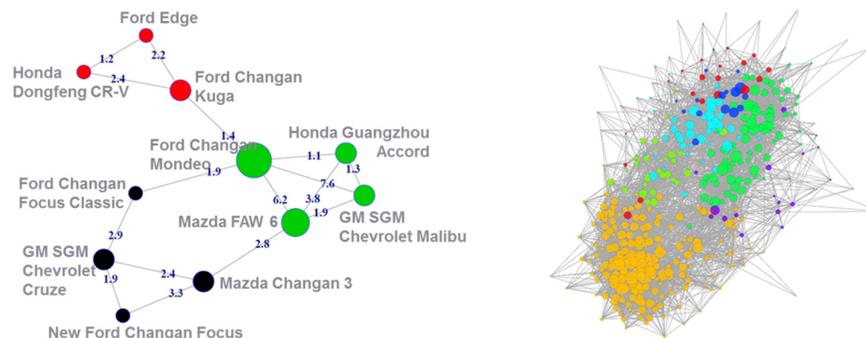


Figure 2. Product association network: only links with lift greater than 1 are shown; left: illustrative network of vehicle associations; right: NCBS network of vehicle associations; seven network communities are found and depicted using different colours

Figure 2 (right) displays the vehicle association network involving all 389 vehicles in the data, visualized by the Fruchterman-Reingold force-directed algorithm. As detailed in our previous work, the size of network nodes represents the degree centrality of nodes (range of connections), which can be used to study the impact on the sales volume data [Wang et al. 2015]. The colors of network nodes highlight the communities (clusters) of network nodes (vehicle models), extracted by the fast-greedy algorithm by [Clauset et al. 2004]. Network community has the property that nodes are strongly connected within one community and less densely connected with the remainder of the network. The community of network nodes suggests the aggregated consideration sets by customers because the frequently mentioned vehicles are grouped together. It is noted that there are high correlations between the vehicle communities and the vehicle 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, etc.).

However, the associations of vehicles in a community is not simply evoked by a single factor (e.g., vehicle segment), but more likely is a consequence of a mixed factors. 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 green community. In the following section, we use JCA 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 et al. 2006] to demonstrate the JCA approach. Assume that we have x_1, \dots, x_q categorical variables (attributes) such as vehicle model and income level on N customer observations. Each x_j is associated with categorical levels $1, \dots, L_j$. 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 $J = L_1 + \dots + L_q$ is the total number of categorical levels for all input variables x .

An example of the indicator matrix \mathbf{Z} with 5 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. Because the indicator matrix \mathbf{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 $\mathbf{B} = \mathbf{Z}'\mathbf{Z}$ which is defined as the cross-tabulation of all categorical levels. Given the Burt matrix \mathbf{B} , the coordinates of columns with respect to the principal axis can be computed by Singular Value Decomposition (SVD) [Greenacre et al. 2006], and the corrected inertia by JCA can be then obtained by iterative updating the solution [Greenacre 2007]. Due to the space limits, technical details are omitted in this paper.

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

	Vehicle model x1			Income level x2		
	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

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 plot is shown in Figure 3.

As noted, the output plot 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 Figure 3(right), 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 Figure 3(right), 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 Figure 3(right), 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.

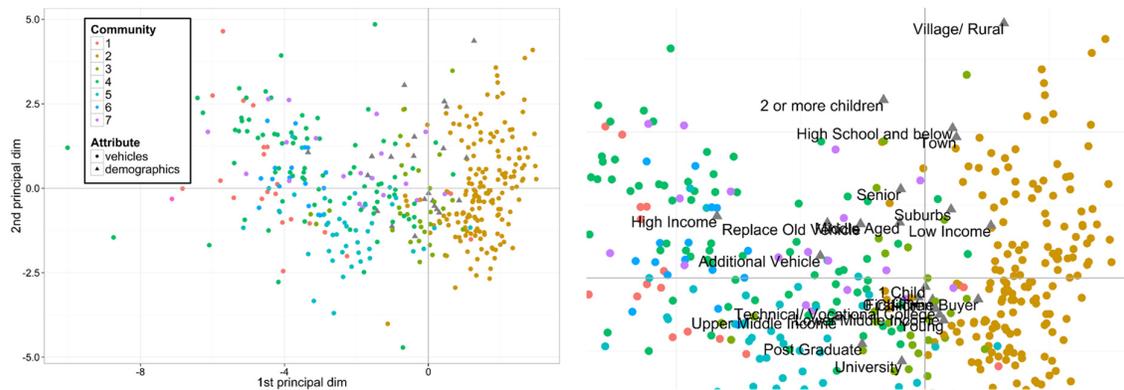


Figure 3. JCA plot of vehicles (in dots) and customer demographics (in triangles) in the first two principal dimensions; vehicle are coloured by network communities; left: full plot; right: enlarged partial plot

In addition to studying the relationships between the set of interested variables, 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 plot, and what attributes explain the formation of the vehicle communities. Figure 3 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”, and “high school and below”. In contrast, the import SUV community (#4, green) is characterized by a completely different set of demographics, including “high income”, “additional vehicle”, and “replace old vehicle”. Both observations are consistent to 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 should stimulate 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 impacts on vehicle associations. We repeat the JCA procedures and display the generated plot in Figure 4. A numerical breakdown of the analysis shows that the first two dimensions in Figure 4(left) explain 77.4% of the total data variance.

Figure 4 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 will also reveal how customers evaluate vehicles differently based on the subjective feelings and what characteristics they care most for each vehicle community. From Figure 4(right), the characteristics of vehicles are relatively more crowded around the center compared to the customer attributes shown in Figure 4(right). 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 Figure 4 (right). 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” and “prestigious”.

Next we use vehicle models and vehicle attributes as explanatory variables to study their impact on vehicle associations. In Figure 5, 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.

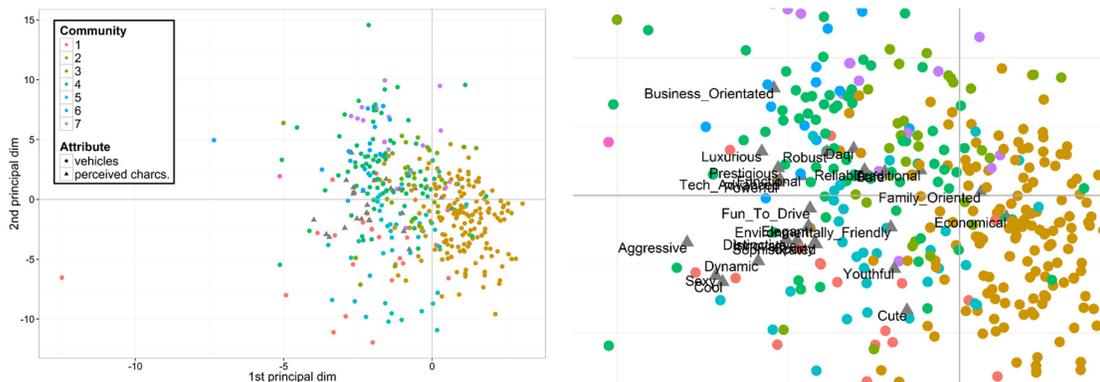


Figure 4. JCA plot of vehicles (in dots) and customer perceived vehicle characteristics (in triangles) in the first two principal dimensions; left: full plot; right: enlarged partial plot

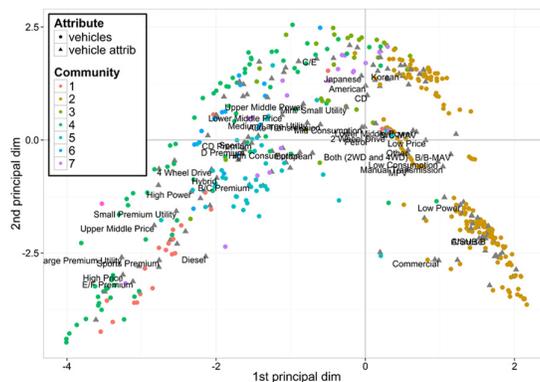


Figure 5. JCA plot of vehicles (in dots) and vehicle attributes (in triangles) in the first two principal dimensions; labels for vehicle brands and origins are hidden for simplicity

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 Figure 6 where the first two dimensions together explain 19.0% of the total variance in the data. It is noted that Figure 6 is more similar to Figure 5 than Figure 3 and Figure 4 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 Figure 6, 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, 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 how people make preference decisions.

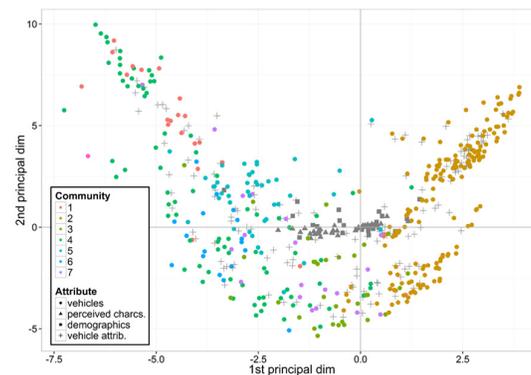


Figure 6. JCA plot of vehicles, customer demographics, perceived vehicle characteristics, and vehicle attributes; different sets of variables are depicted in different shapes; communities are shown in colours

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 visual representations, in the next section, our interest is to further quantify the importance of the multiple factors whose similarities (differences) may influence the formation of consideration sets. The results of JCA will be integrated to create predictive models for assessing the association strength of vehicles in the product network. Such model can also be used to predict the change of consideration decisions as a result of the change of product configurations.

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/customer attributes, and then use them to predict the network structure of co-consideration relationships. As shown in Figure 7, 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 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).

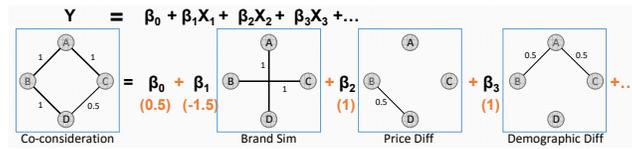


Figure 7. Illustration of MRQAP

The MRQAP analysis is employed to overcome the limitation of the independence assumption in traditional Ordinary Least Square (OLS) regression. 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, where the data matrix is transformed corresponding to a permutation of rows and columns, with rows and columns being permuted in the same way. OLS coefficients are then estimated from the permuted matrices. 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 FLSF 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 Figure 7, 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 difference between the purchase prices of the two vehicles. In the “demographic difference” network, we use the first two principal coordinates derived from JCA in Figure 3 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:

$$X_{ij} = I\{x_i = x_j\} \quad (1)$$

where $I\{\cdot\}$ represents the indicator function; 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:

$$X_{ij} = |x_i - x_j| \quad (2)$$

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:

$$X_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2 \quad (3)$$

where $\|\cdot\|_2$ is the L2-norm; \mathbf{x}_i and \mathbf{x}_j are the first two dimensions of principal coordinates 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]:

$$Y_{ij} = \log(1 + lift_{ij}) \quad (4)$$

6.2 Results

The result of MRQAP based on the NCBS data is reported in Table 2. We compare three MRQAP models to highlight the contributions of different types of variables.

**Table 2. 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.						

Model 1 formulates a MRQAP model that regresses the lifts of vehicle associations on the similarity (or difference) of vehicle attributes only. These attributes are mostly engineering attributes 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 co-considerations. Consistent to the results from JCA, vehicles sharing similar attributes are more likely to be co-considered, as explained by the positive coefficient for a similarity network, and the negative coefficient for a difference network. By examining the values of the coefficients, we see that the vehicle *segment* has the strongest effect, meaning that the structure of the *segment* similarity network is the most closely related one 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 model the characteristics network defined by the coordinates of vehicles in JCA shown in Figure 4. This time, we predict the co-considerations 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 differences of the perceived characteristics have little influence on co-consideration relationships compared to the influence from 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 Figure 3, 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

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 survey 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, pp. 145-154.
- Benzécri, J.-P., "L'analyse des données", *Vol.2, Paris, Dunod*, 1973.
- Clauset, A., Newman, M. E., Moore, C., "Finding community structure in very large networks", *Physical review E*, *Vol.70, No.6*, 2004.
- Dekker, D., Krackhardt, D., Snijders, T. A., "Sensitivity of MRQAP tests to collinearity and autocorrelation conditions", *Psychometrika*, *Vol.72, No.4*, 2007, pp. 563-581.
- Dieckmann, A., Dippold, K., Dietrich, H., "Compensatory versus noncompensatory models for predicting consumer preferences", *Judgment and Decision Making*, *Vol.4, No.3*, 2009, pp. 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, pp. 67-83.
- 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., Toubia, O., Evgeniou, T., Befurt, R., Dzyabura, D., "Disjunctions of conjunctions, cognitive simplicity, and consideration sets", *Journal of Marketing Research*, *Vol.47, No.3*, 2010, pp. 485-496.
- Hauser, J. R., Wernerfelt, B., "An Evaluation Cost Model of Consideration Sets", *Journal of Consumer Research*, *Vol.16, No.4*, 1990, pp. 393-408.
- Henderson, G. R., Iacobucci, D., Calder, B. J., "Brand diagnostics: Mapping branding effects using consumer associative networks", *European Journal of Operational Research*, *Vol.111, No.2*, 1998, pp. 306-327.
- Krackhardt, D., "Predicting with networks: Nonparametric multiple regression analysis of dyadic data", *Social networks*, *Vol.10, No.4*, 1988, pp. 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*, *Vol.136, No.3*, 2014.
- Netzer, O., Feldman, R., Goldenberg, J., Fresko, M., "Mine your own business: Market-structure surveillance through text mining", *Marketing Science*, *Vol.31, No.3*, 2012, pp. 521-543.
- Sosa, M. E., Eppinger, S. D., Rowles, C. M., "A network approach to define modularity of components in complex products", *Journal of Mechanical Design*, *Vol.129, No.11*, 2007, pp. 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)*, 2015a.
- 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*, 2015b.
- Wasserman, S., Faust, K., "Social network analysis: Methods and applications", *Vol.8, Cambridge university press*, 1994.

Dr. Wei Chen, Professor
Northwestern University, Department of Mechanical Engineering
2145 Sheridan Road, Tech B224, IL 60208-3111 Evanston, United States
Email: weichen@northwestern.edu