

MODELING CUSTOMER PREFERENCES USING MULTIDIMENSIONAL NETWORK ANALYSIS IN ENGINEERING DESIGN

Mingxian Wang, Wei Chen¹

Department of Mechanical Engineering, Northwestern University, Evanston, IL

Yun Huang, Noshir S. Contractor

Science of Networks in Communities, Northwestern University, Evanston, IL

Yan Fu

Ford Motor Company, Dearborn, MI

ABSTRACT

Motivated by overcoming the existing utility-based choice modeling approaches, we present a novel conceptual framework of multidimensional network analysis (MNA) for modeling customer preferences in supporting design decisions. In the proposed Multidimensional Customer-Product Network (MCPN), customer-product interactions are viewed as a socio-technical system where separate entities of “customers” and “products” are simultaneously modeled as two layers of a network, and multiple types of relations, such as consideration and purchase, product associations, and customer social interactions, are considered. We first introduce a unidimensional network where aggregated customer preferences and product similarities are analyzed to inform designers about the implied product competitions and market segments. We then extend the network to a multidimensional structure where customer social interactions are introduced for evaluating social influence on heterogeneous product preferences. Beyond the traditional descriptive analysis used in network analysis, we employ the Exponential Random Graph Model (ERGM) as a unified statistical inference framework to interpret complex preference decisions. Our approach broadens the traditional utility-based logit models by considering dependency among complex customer-product relations, including the similarity of associated products, “irrationality” of customers induced by social influence, nested multi-choice decisions, and correlated attributes of customers and products.

1. INTRODUCTION

Understanding customer preferences, interests, and needs is critically important in developing successful products (Ulrich, 2003). In the past decade, the theory of decision-based design and various preference modeling techniques have been continuously developed for this purpose (Shiau and Michalek, 2009, Frischknecht et al., 2010, Morrow et al., 2014, Chen et al., 2013). However, analytical modeling of customer preferences in product design is inherently

¹ Corresponding Author, weichen@northwestern.edu, Wilson-Cook Professor in Engineering Design

difficult as it faces challenges in modeling heterogeneous human behaviors, complex human interactions, and a large variety of product offerings. Motivated by overcoming the existing utility-based choice models, we propose a novel multidimensional network analysis (MNA) approach, rooted in social network analysis for modeling complex customer-product relations. The focus of this paper is on presenting the conceptual framework of the proposed Multidimensional Customer-Product Network (MCPN) together with the development of Exponential Random Graph Model (ERGM) for preference data analysis. By using the network approaches developed in this paper, designers could, in principle, gain a better insight into the customers' desires and preferences, the market structures, the competitive landscape, the strength of social influence, and the key attributes of their own and their competitors' products. In new product development and future strategic planning, the constructed network model can be used to identify the right product configurations for targeted group of customers in a new market.

In recent years, disaggregate quantitative models such as Discrete Choice Analysis (DCA) (Ben-Akiva and Lerman, 1985, Train, 2009) have been widely studied by the engineering design research community for consumer preference modeling. Following the random utility theory, the customer purchase decision is captured by a utility function of product attributes/features and customer attributes (e.g., social demographic and usage attributes) (Hoyle et al., 2010, He et al., 2012). Even though DCA provides a probabilistic approach for modeling customer heterogeneity, there are several major obstacles regarding their use in practical design applications:

- **Dependency of Alternatives.** Standard logit models usually ignore correlations in unobserved factors over product alternatives by assuming observations are independent, i.e., whether a customer chooses one product is not influenced by adding or substituting another product in the choice set. This is also known as the Independence of Irrelevant Alternatives (IIA) property, whose implication is not realistic for applications with similar product offerings. Though advanced logit models have been developed to address this issue by introducing certain correlation structures among the error terms, they cannot accommodate any dependent decisions explicitly.
- **Rationality of Customers.** The utility-function based choice modeling approach assumes customers make rational and independent decisions. However, in reality customers influence each other, and their socially influenced decisions can sometimes be considered “irrational.” As such, it is reasonable to expect that social effects, such as geographical proximity, communication ties, friendship connections, and social conformity have large influences on customer attitude and behavior.
- **Correlation of Decisions.** Correlated decisions often exist, such as in forming a consideration set. It is important to realize that decisions in such situation are often nested within one another. For example, the decision of how many products and what products to consider could be nested. Unfortunately, classical regression models ignore these correlations, and therefore, cannot estimate the influence of correlated decisions.
- **Collinearity of Attributes.** To evaluate the underlying preference for each product attribute (feature), it is often desirable that preference data used for modeling has little to no collinearity. However, revealed preference data is very vulnerable to collinearity as the data is drawn directly from the real market. For example, low price vehicles are more possible to have smaller engine capacity and as a result, low fuel consumptions. However, it is hard to tell whether customers are buying cars because they are low price or because they are fuel efficient. The presence

of collinearity implies that the contribution of each attribute is difficult to evaluate separately using utility-based logit models.

Our goal in this work is to overcome the limitations of existing quantitative methods for modeling customer preferences in engineering design. We aim to develop a preference model that broadens the utility-based DCA by considering complex customer-product relations, including the similarity of associated products, “irrationality” of customers induced by social influence, nested multi-choice decisions, and correlated attributes of customers and products. To this end, we propose a novel Multidimensional Customer-Product Network (MCPN) framework as shown in Fig. 1. As seen, customer-product interactions form a complex socio-technical system (Trist, 1981), not only because there are complex relations between the customers (e.g., social interactions) and amongst the products (e.g., market segmentation or product family), but also because there exist multiple types of relations between customers and products (e.g., “consideration” versus “purchase”). Our research premise is that, similar to other complex systems exhibiting dynamic, uncertain, and emerging behaviors, customer-product relations should be viewed as a complex socio-technical system and analyzed using social network theory and techniques. The structure and topological characteristics identified in customer-product networks can reveal emerging patterns of customer-product relations while taking into account the heterogeneities among customers and products.

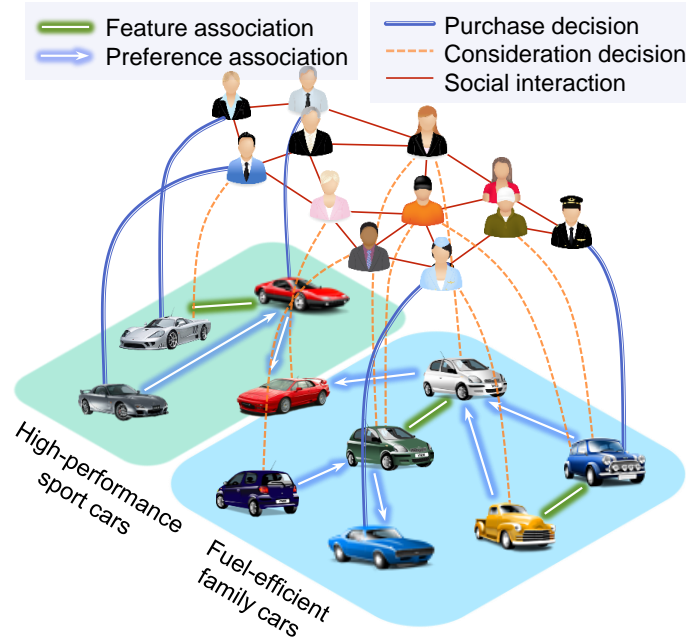


Figure 1: Customer-Product Relations as a Complex Network System, with Five Types of Relations and Two Types of Nodes

In literature, network analysis has emerged as a key method for analyzing complex systems in a wide variety of scientific, social, and engineering domains (Wasserman and Faust, 1994). The approach provides visualization of complex relationships depicted in a network graph, where *nodes* represent individual members and *ties/links* represent relationships between members. Built upon conventional network analysis, social network analysis views social relations in terms of network theory, and the links in the observed network are explained by the underlying social

processes such as self-interest, collective action, social exchange, balance, homophily, contagion, and co-evolution (Monge and Contractor, 2003).

Most existing applications of network analysis are *unimodal* or *unidimensional* that contain a single class of nodes (either human or non-human artifact) and a single type of relation. For example, our previous research characterizes customer consideration preferences over a decision set of products through a unidimensional product association network, where the link represents the customer consideration decision and the node represents the product offering. Even though the analysis of unidimensional network structures have been used in our early research to understand product competitions (Wang et al., 2015), and predicting heterogeneous choice sets using Discrete Choice Analysis (Wang and Chen, 2015), the preference analysis can only be performed in the aggregated level representing the average preference decision across the customer population. A more complex structure is the bipartite network, which contains both human and non-human technological elements as nodes, and a single type of relation connecting the two sets of nodes. With the addition of the second types of node (consumer) into the network, researchers can model preferences at the disaggregated individual level as opposed to the aggregated group level preferences studied in our previous work (Wang et al., 2015, Wang and Chen, 2015). In literature, recent social network researchers put more emphasis on the development of *multidimensional* networks (Contractor et al., 2011), which include multiple types of nodes, as well as multiple types of relations represented by non-directed or directed links at multiple levels. It has been recognized that the multidimensional structures can be useful in studying how technologies can simultaneously shape and be shaped by the social structures into which they are introduced, because technology and people are modeled in two separate layers of a network. On one hand, the social structures can influence how people conceive a new technology (or a product), as well as whether and how they will use it (Rice and Aydin, 1991, Karahanna et al., 1999, Kraut et al., 1998). On the other hand, new technologies (or products) could bring changes to social and communication relationships among people (DiMaggio et al., 2001). To the authors' knowledge, this paper represents the first attempt to introduce multidimensional network analysis into engineering design. The complex customer-product interactions are represented as a multidimensional network where multiple relationships are considered, including social network relations among customers, association relations among products, as well as preference relations between customers and products.

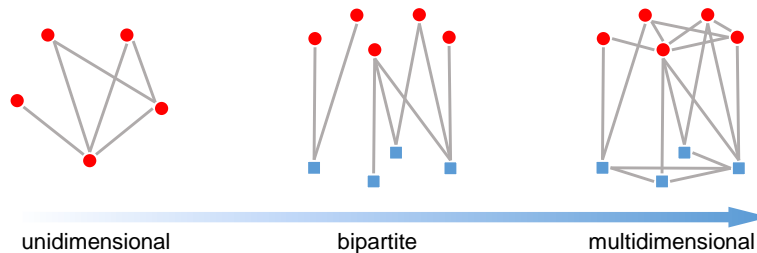


Figure 2: Development of Network Structures

Beyond most existing network analyses that are descriptive in nature, our research introduces the *Exponential Random Graph Model (ERGM)* as a unified statistical inference framework for MNA. ERGM is increasingly recognized as one of the central approaches in analyzing social networks (Lusher et al., 2012, Robins et al., 2007, Wang et al., 2013). ERGMs account for the presence (and absence) of network links and thus provide a model for

analyzing and predicting network structures. ERGMs have several advantages over the utility-based logit models: (1) Network links are modeled to be interdependent in ERGM rather than assumed to be independent; (2) ERGMs can incorporate binary, categorical, and continuous node attributes to determine whether they are associated with the formation of network links, (3) ERGMs are capable of characterizing local and global network features; (4) ERGMs can be applied in flexible ways to many different types of network and relational data; (5) Data used for fitting ERGMs can be cross-sectional or longitudinal (change with time), and a dynamic model can be built to study the emergence and dynamics of a network; (6) In contrast to a machine learning model that focuses on prediction, ERGM is an explanatory model whose results can be used to derive behavioral theories and design implications.

This paper employs MNA for the study of customer-product relations as a complex Multidimensional Customer-Product Network (MCPN) in the context of engineering design. While the proposed MCPN is expected to be widely applicable for characterizing any type(s) of preference relations (e.g., consideration decision, purchase decision), the detailed development in this paper is devoted to modeling consideration decisions among product alternatives. Our emphasis in this paper is on demonstrating the uniqueness and potential of the MNA approach rather than testing the model prediction capability *per se*. Examining how to improve the prediction accuracy of such network models to improve the quality of design decisions belongs to future work. The rest of the paper is organized as follows. Sec. 2 introduces the technical background and recent accomplishments in social network research. Sec. 3 describes the conceptual framework of the proposed approach and illustrate the development of MCPN progressively from a unidimensional structure to a multidimensional structure with multiple types of nodes and links. Sec. 4 develops two network implementations using the vehicle preference data in China market -- a descriptive approach for a unidimensional network and a statistical inferential technique for a multidimensional network. Finally, Sec. 5 presents the pros and cons of the MNA approach and the opportunities for future research.

2. TECHNICAL BACKGROUND

2.1 Network analysis in product design and market study

Network analysis has attracted considerable interest in product design and market study. In product design, network analysis has been used to characterize a complex product as a network of components that share technical interfaces or connections. Using the network metrics such as “centrality”, Sosa et al. (Sosa et al., 2007) defined three measures of modularity as a way to improve the understanding of product architecture. Based on Sosa’s work, Fan et al. (Fan et al., 2013) developed a bottom-up strategy for modular product platform planning. A recent work by Sosa (Sosa et al., 2011) found that proactively managing the use of network structure (such as hubs) may help improve the quality of complex product designs. Network analysis has also been applied to studying designers’ network for understanding organizational behavior (Contractor et al., 2011) and improving multidisciplinary design efficiency (Cormier et al., 2012). In market study, text-mining apparatus has been integrated into a network analysis framework to understand customers’ top-of-mind associative network of products based on a large-scale, customer generated dataset on the Web (Netzer et al., 2012). However, all the aforementioned product/feature networks are unidimensional, without including customers and their (preferences) relations to products in the same network. In contrast to the existing unidimensional product network analysis approaches, our multidimensional customer-product network (MCPN) is built with both product and customer nodes, together with product feature associations and

customer social network, to understand how customer decision-making interacts with product attributes and how social influence affects individual decisions for new products.

2.2. Modeling the impact of social influence

There is growing recognition that modeling the impact of social influence is important in product design (Aral and Walker, 2011). A comprehensive study of how peer influence affects product attribute preference was provided by Narayan et al. (Narayan et al., 2011) who modeled three different mechanisms of social influence. By combining traditional conjoint analysis on product features with peer influence, their work showed that peer influence causes people to change perspective on product importance, and that some product attributes are more sensitive to change than others. However, the approach requires a strict format of survey data to evaluate the attitude change before and after exposure to peer influence.

In modeling social influence in customer vehicle choices, a simulation-based approach has been developed in our earlier research to capture the dynamic influence from social networks on the adoption of hybrid electric vehicles (He et al., 2014). The social network impact is captured via introducing “social influence attributes” into the discrete choice utility function. The effects of these attributes are assessed through the social network simulation, where the network was constructed based on the “social distances” measured by the dissimilarities of customers’ social profiles. This approach was demonstrated through a vehicle case study where a customer’s decision in choosing an eco-friendly alternative fuel vehicle could be influenced by neighbors and friends modeled as a small world network. Similar treatments of using a small world network for capturing social influence have also been found in (Delre et al., 2007, Watts and Strogatz, 1998). In this research, a multidimensional network approach is proposed to measure simultaneously customer-customer social interactions together with customer-product preference relations for assessing social impact on preference decisions. A simulation-based social network construction approach, similar to (He et al., 2014), is applied to convert customer attribute vectors into relational data in constructing the social network as a part of the MCPN which takes into account the interdependence of attributes and the interactions between customers and products.

2.3. Advances in Social Network Analysis

In the past decade, social network scholarship has made a concerted effort to move from describing a network to developing techniques that explain the emergence and dynamics of networks. Development of analytical techniques to explain the emergence of networks is often motivated by the *multitheoretical multilevel (MTML)* framework (Monge and Contractor, 2003). Social network models are multi-theoretical because of the growing recognition among social networks researchers that the emergence of a network can rarely be adequately explained by a single theory. Therefore, social network models combine disparate theoretical generative mechanisms, such as self-interest, collective action, social exchange, balance, homophily, proximity, contagion, and co-evolution. Social network models have multilevel interpretations because the emergence of a network can be influenced, for instance, by theories of *self-interest* that refer to characteristics of actors (at the individual level), theories of *social exchange* that describe links between pairs of actors (at the dyadic level), theories of *balance* that explain the configuration of links among three actors (at the triadic level), and theories of *collective action* that explain configurations among larger aggregates of actors (at the group or network level).

Among the network modeling techniques, the Exponential Random Graph Model (ERGM) provides the statistical inference framework for multidimensional network analysis (MNA). Technically, we can define matrix \mathbf{Y} as a random graph in which rows and columns represent customers and products, respectively. $Y_{ij} = 1$ refers to a relation, such as the preference (consideration or purchase) decision between customer i and product j , and 0 otherwise. ERGMs have the following form:

$$P_{\theta}(\mathbf{Y} = \mathbf{y}) = \frac{1}{c(\boldsymbol{\theta})} \exp\{\boldsymbol{\theta}^T \mathbf{z}(\mathbf{y})\}, \quad (1)$$

where (i) \mathbf{y} is the observed network, a random realization of \mathbf{Y} ; (ii) $\mathbf{z}(\mathbf{y})$ is a vector of network statistics corresponding to network characteristics in \mathbf{y} , and the settings of product and consumer attributes; (iii) $\boldsymbol{\theta}$ is a parameter vector indicating the effects of the network statistics; (iv) c is the normalizing constant that ensures the equation is a proper probability distribution. Eqn. (1) suggests that the probability of observing any particular graph (e.g. MCPN) is proportional to the exponent of a weighted combination of network characteristics: one statistic $\mathbf{z}(\mathbf{y})$ is more likely to occur if the corresponding θ is positive. Examples of possible $\mathbf{z}(\mathbf{y})$ statistics used in MCPN is detailed in Sec. 3.3. Our research aims to interpret the meaning of these network effects $\mathbf{z}(\mathbf{y})$ in order to understand customer-product relations for product design.

3. A MULTIDIMENSIONAL NETWORK APPROACH FOR PREFERENCE MODELING

3.1. The Multidimensional Customer-Product Network (MCPN) Framework

In this paper, we recast the problem of modeling customer preferences as network modeling of customer-product relations. We view engineering products as an inherent part of the expanded social network along with human actors. Fig. 3 describes the structure of the MCPN framework, which is characterized by two classes of nodes at two layers (“product” and “customer”) and multiple types of relations within and between the two layers.

The **product layer** contains a collection of engineering products \mathbf{P} (e.g., vehicles, electronics and appliances, software). Products are connected by various links which can be either directed or non-directed. *Directed links* often involve product hierarchy or preference, while *non-directed links* imply product similarity or association. Product attributes or features, quantitative (e.g. fuel efficiency, horsepower) or qualitative (e.g. safety, styling), can be taken into account as nodal attributes. Similar attributes/features between products are represented as association links in the product network. Alternatively, product associations can be identified by their co-consideration relations from customers. The **customer layer** describes a social network consisting of a customer population \mathbf{C} who make decisions or take actions. Each customer has a unique profile (e.g., socioeconomic and anthropometric attributes, purchase history and usage context attributes, etc.) which potentially affects customer preference decisions. Links between two customers represent their social relations, such as friendship or communication. The structural tendencies of these social relations reflect the underlying social processes for creating and maintaining links such as homophily and proximity (Monge and Contractor, 2003). Much of the developed literature in social network analysis can be employed here to construct a customer social network. **Customer-product relations** are indicated by various human activities such as *purchase* and *consideration* decisions. The *customer-product links* are created between two sets of nodes from two adjacent layers, representing customer preferences. As shown in Figure 3, if a customer purchases a product, there is a solid line linking the customer and product nodes. If a customer considers a product, the link between the two

nodes is marked as a dashed line. As noted, a customer can consider several products at the same time while the final purchase is only one or none. These preference links can be flexibly constructed by various sources of data, e.g., survey data, transaction data, and user-generated text data.

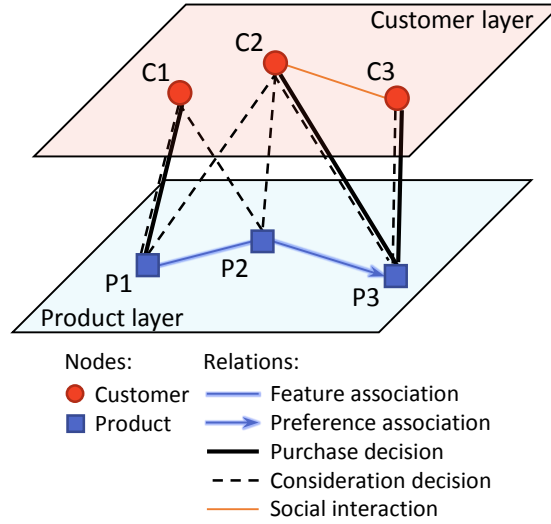


Figure 3: Multidimensional Customer-Product Network

As seen, the proposed MCPN framework can capture rich information on dependency in a complex socio-technical system so as to assist product design decision-making. A combined analysis of all relations mentioned above allows designers to evaluate product decisions not in isolation, but with expectation that the market system will react to the planned decisions, and any design change may easily affect other connected entities across the network in ways that were initially unintended.

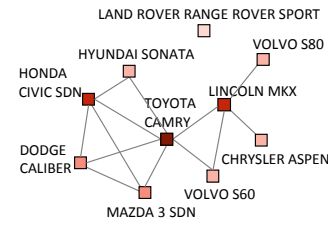
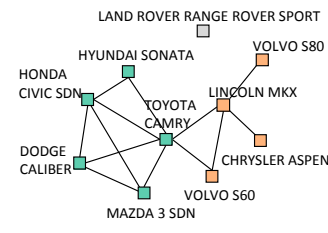
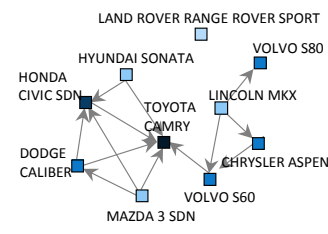
3.2. Unidimensional Network Analysis of Product Associations

Our development of MCPN started with the unidimensional network analysis to a single layer network with only product nodes and associations. The unidimensional network can be viewed as a compressed but simplified version of the more complicated bipartite (customer-product) networks by projecting it to a single layer (Wasserman and Faust, 1994). The unidimensional network enables designers to explore the use of descriptive metrics in identifying aggregated product associations that can reveal the implied product similarity and diversity, product market competence, product market segmentation, and other opportunities for design improvements.

The links in a product association network can be constructed in many ways. For example, using the customer preference data, a *customer-driven product association network* can be established, where the links between products reflect the proximity or similarity of two products in customers' perceptual space. Standard measures of association rules, such as the "lift", can be used to quantify the strength of the connection (Tan et al., 2004) between two products based on how often they are in the same consideration set. Alternatively, a *feature-driven product association network* can be established with the help of product specification data, where the association between products can be determined by measuring the similarity of product attributes/features from designers' point of view. Distance measures commonly used in content-based recommender systems, such as Jaccard index (Real and Vargas, 1996), Cosine

similarity (Chowdhury, 2010), Gower similarity ((Gower, 1971)) and etc., can be applied to assess the strength of connections.

Table 1: Examples of Descriptive Network Analysis for Analyzing Customer-Driven Product Associations

| Network Analysis | | Solution Techniques | Network Topology |
|------------------|---|--|---|
| Centrality | Centrality involves the identification of the ‘most competitive’ products in the network (Wasserman and Faust, 1994). We assume that more central (or more connected) products have higher levels of survivability in market competitions as a result of its structural advantages. | Measuring centrality can be based on various properties of a node, e.g., number of direct connections to all other nodes (degree), minimum distance to all other nodes (closeness), and maximum occurrence on the path of two other nodes (betweenness) (Wasserman and Faust, 1994). |  <p>Figure 4a. Vehicle Centrality in Network, Constructed Based on Co-consideration Data</p> |
| | Community refers to the occurrence of groups of nodes that are more densely connected internally than with the rest of the network (Newman and Girvan, 2004). If appropriate communities are detected, the network can be collapsed into a simpler representation without losing much useful information. | The modularity maximization method (Newman and Girvan, 2004) can be used as the objective function to capture the quality of a network structure. The problem is solved as an NP-hard optimization problem. |  <p>Figure 4b. Vehicle Community in Network, Constructed Based on Co-consideration Data</p> |
| | Hierarchy is formally defined as a strict partially ordered set which can be represented as a directed network (Corominas-Murtra et al., 2013), where each element of the set is a node and the partial ordering ($P1 < P2$) gives an edge from P1 to P2. The directed link reflects customers’ aggregated preference across the population. | To find local hierarchies of nodes, centrality metrics can be applied as well to a directed network configuration. To bring global order to the nodes, heuristic search algorithms, e.g. Google’s PageRank (Page et al., 1999), can be employed to find the best hierarchy in a polynomial time. |  <p>Figure 4c. Vehicle Hierarchy in Network, Constructed Based on Co-consideration and Purchase Data</p> |

The descriptive network analysis involves the computation of topological measures to assess the position of nodes and the implication of structural advantages. Examples for analyzing customer-driven product associations are provided in Table 1. **Centrality** (Freeman, 1979, Wasserman and Faust, 1994) measures a product’s competitiveness, indicated by its level of connectivity to other products. As an example in Fig. 4a, a vehicle network is constructed based on co-consideration data, indicating if two vehicles are co-considered by the majority of customers. As seen, Toyota Camry is more “central” to other vehicles, implying it has the potential to satisfy a broader range of customers. Honda Civic SDN and Lincoln MKX are the next widely-considered cars in customers’ minds. Network **community** (Clauset et al., 2004, Newman and Girvan, 2004) analysis identifies products in the same community based on the link strengths and connections. In Fig. 4b, two distinct communities (“compact vehicles” in green dots and “high-performance midsize vehicles” in orange dots) are found in the vehicle co-consideration network. The emergent

product communities can be used to detect consumer choice set and potential product competitions (Wang and Chen, 2015). “Crossover” vehicles that belong to multiple (overlapping) communities can also be identified through this analysis (Gregory, 2007, Palla et al., 2005). Network **hierarchy** (Corominas-Murtra et al., 2013, De Vries, 1998, Gupte et al., 2011) is illustrated by the directional network links in Fig. 4c which encode preference ranking based on both consideration and purchase data. Products with high preference ranks (e.g., Camry with many incoming edges) are shown in darker colors. Note that in analyzing feature-driven product association networks where products are linked based on shared features, centrality, community, and hierarchy have different implications. For example, community implies product families where common features are shared among products.

Although the unidimensional network approach can describe interdependencies in relational data, the method cannot provide quantitative assessment of the impact of product attributes for a particular group of customers of interest. Further, the unidimensional network analysis studies customers’ averaged (aggregated) preference across the population. Advanced network modeling approaches that capture disaggregated preference behaviors of individual customers are needed as examined next.

3.3. Analyzing Multidimensional Network Considering Product Associations

To model heterogeneous customer preferences in products with close associations, we integrate the product association links with customer-product preference relations as a multidimensional network (see Fig. 5), including two classes of nodes (customers and products), multiple types of customer preference relations, and association relations among products. By introducing the information from the second mode (i.e. customers), we aim to develop a network model capable of capturing customer preference heterogeneity and multiple dependent decisions, while considering product feature associations.

Beyond existing network approaches that are mostly descriptive in nature, we use ERGM as a unified statistical framework to analyze the MCPN. In ERGMs, the observed network is considered one realization of an underlying probabilistic distribution, without assuming the independence of nodes or links. A local topological configuration in the network, i.e. a set of connected nodes and links, is regarded as an exploratory variable representing the structural features of potential interest. Networks in the distribution are assumed to be “built up” from the localized patterns represented by the structural features. ERGM literature has established more than 20 different types of effects (Lusher et al., 2012) for describing the various forms of dependence that exist in the relational data within social networks. Examples of effects, their configurations, and interpretations are provided in Table 2. Our focus is to interpret the meaning of these effects to understand customer-product relations for product design.

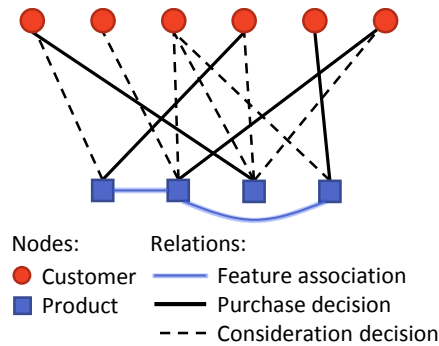


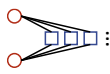



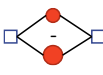




Figure 5: Multidimensional Network Considering Product Associations

The network effects fall into three categories: *pure structural effects*, *attribute-relation effects* involving product/customer attributes, and *cross-level effects* involving both between-level and within-level relations. *Pure structural effects* are related to the well-known structural regularities in the network literature (e.g., effects [A] to [C] in Table 2); *attribute-relation effects* assume the attributes of products/customers can also influence possible tie formations in a given structure. At the two-node level (effects [A], [D], [E]), interpretation resembles the attribute-effect in a logistic regression [121, 122]. The main effect (effect [D]) can be used to test how attractive a product attribute is. The interaction effect (effect [E]) captures whether certain features are favored by a particular group of customers or not. Beyond conventional logistic models, the network approach also evaluates higher-order effects such as at the levels of three-node (effect [F]) and four-node (effect [G]). The product association relations can be captured by the *cross-level effects* (effect [H]) that integrate customer preferences with product similarities. In this way, the analysis can explain whether certain types of customers tend to consider product alternatives associated with a specific set of attributes.

Table 2: Examples of Network Effects in MCPN, with Graphical Configurations and Design Interpretations.

| Pure Structural Effects | Configuration | Interpretation |
|--|---|--|
| [A] Density |  | This effect captures the baseline propensity of forming a link. It is similar to the intercept in a regression model. |
| [B] Alternating k-stars for products |  | This effect measures the dispersion of the degree distribution. Alternatively, it can be thought as a test of the “rich get richer effect”. Example: A positive parameter indicates that the network links are centralized around a few high-degree nodes of products. |
| [C] Alternating k-cycles for customers |  | This effect captures the propensity of customers to engage in closed structures. Example: Two customers considered the same product also consider some other products together. |
| Attribute-Relation Effects | Configuration | Interpretation |
| [D] Main effect |  | This effect captures whether the binary attribute or higher scores on a continuous attribute tend to express more links. Example: A significant negative parameter for vehicle fuel consumption means fuel efficient cars are more likely to be considered by customers. |
| [E] Interaction effect |  | This effect captures the interaction of the nodes between different types. Example: A significant positive coefficient for family size of customers and vehicle size of products suggests customers from large families tend to consider large size cars. |
| [F] Two-paths difference effect |  | This effect captures the differences of continuous attributes on the ends of a 2-path. Example: For “price”, a negative significant estimate suggests that two cars with little price differences tend to be considered together by the same customer. |
| [G] Four-cycle difference effect |  | This effect captures whether closed structure is more likely to occur involving two customer nodes with similar attributes. Example: For “income”, a negative significant coefficient means that customers with similar income tend to share many cars in consideration. |
| Cross-level Effects | Configuration | Interpretation |
| [H] Association based closure effect |  | This effect captures whether a closed structure is more likely to occur involving two product nodes with an association link. Example: A negative significant coefficient means that customers do not tend to consider two cars with many common features at the same time. |
|  | | |

Once the network effects of interest are identified, their significance can be determined by estimating the model parameters of an ERGM via likelihood maximization, given the observed network data. As the exact maximization of the likelihood function requires a summation over all possible configurations of the network and is computationally demanding, approximation techniques (e.g., maximum pseudolikelihood (Frank and Strauss, 1986), Markov Chain Monte Carlo maximum likelihood (Geyer and Thompson, 1992)) can be employed to determine the estimates of effects.

Compared to a unidimensional network (Sec. 3.2), a multidimensional network provides a more natural way to model relations between two different classes of nodes (customers and products) and the non-hierarchical association relations between products. Moreover, its capacity to preserve two types of nodes allows designers to parse out the unique contribution of different types of nodes to the overall network structure. Its ability to integrate product networks and customer-product relations allows designers to model interdependent product relations and correlated preference decisions explicitly, without specifying complicated error structures as often done in DCA.

3.4. Analyzing Multidimensional Network incorporating Social Influence

To account for the effect of social influence on customer preference decisions, we further expand the multidimensional network structure to simultaneously measure within-layer social relations, within-layer product associations, and between-layer customer-product relations (Fig. 6).

The proposed multidimensional network allows the evaluation of both the “peer effect” and the more general “crowd effect” (Urberg, 1992), depending on how product associations and social relations are defined. Relations between customers are used to model “peer effect” on customer attitudes and preferences. The term “peer” has a broad meaning which may include “friends,” “neighbors,” “experts,” “relatives” or even “online reviewers” with whom customers may exchange information about new products. The preference hierarchies among products, as defined in Sec. 3.2, can be used to capture the effect of “social crowd”. The evaluation of social influence is done by assessing the structural tendencies of networks informed by social influence theories (Table 3). Using ERGMs, one can quantify the effects of social influence by statistically estimating the extent to which structural tendencies implied by social theories influence the probabilities of the observed network. Similar to the network effects in Table 2, customer and product attributes can be incorporated into the social influence structures for investigating how social influence varies across customers and products.

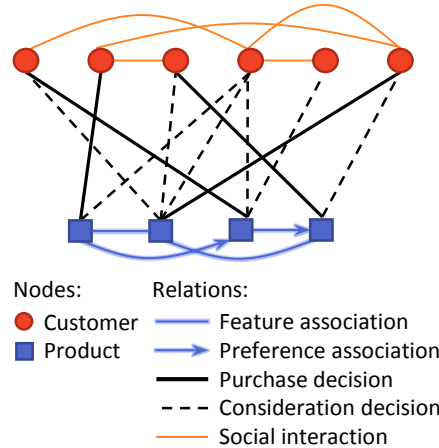
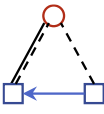
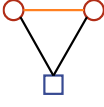


Figure 6: Multidimensional Network Considering Social Interactions

Due to the complexity of data collection, customer social network data is often not collected in consumer surveys. An alternative is to construct social relations through network simulations (He et al., 2014), based on certain hypotheses of network structure and “social distance” measured by the collected customer profiles. For example, based on the theory of homophily (McPherson et al., 2001), we can assume that two nodes with shorter social distance (similar customer attributes) are more likely to be connected. Unlike the prior research that incorporates social influence as customer attributes, this research employs the ERGM to assess the social influence effects. In theory, compared to the use of discrete choice analysis, one should draw more reliable conclusions based on the results from the network approach, because of its capability of handling correlated node attributes and interdependent link relations, which avoids faulty inferences on covariates (Cranmer and Desmarais, 2011).

Table 3: Examples of Social Influence Effects in Multidimensional Network

| Social Influence Effects | Config. | Interpretation |
|--|---|--|
| Crowd effect on purchase |  | When comparing two products under consideration, a customer is more likely to purchase the one favored by the majority of customers. |
| Peer effect on purchase |  | Customers tend to purchase the product that their “peers” recommended, either through use or discussion. |
| <div style="display: flex; justify-content: space-around; align-items: center;"> □ Product w/o attributes ○ Customer w/o attributes </div> | | |

4. CASE STUDY – VEHICLE PREFERENCE MODELING

In this section, two implementations on modeling vehicle preferences in the growing China market are presented to demonstrate the proposed methodology. From simple to complex, our research first examines the use of unimodal networks in Sec. 4.1, studying the product associations from customer’s points of view, and identifying product co-considerations and preference hierarchies. In Sec. 4.2, a multidimensional network is constructed where the ERGM is applied for analyzing customer preferences towards vehicle products, while assessing simultaneously the impact of customer social interactions and product associations. The examples are developed to illustrate the new insights that can be gained but cannot be addressed using the traditional DCA, as well as the flexibility and broad applicability of network analysis to modeling the individual-level preference data.

4.1. Using Unidimensional Network for Modeling Vehicle Associations and Hierarchies

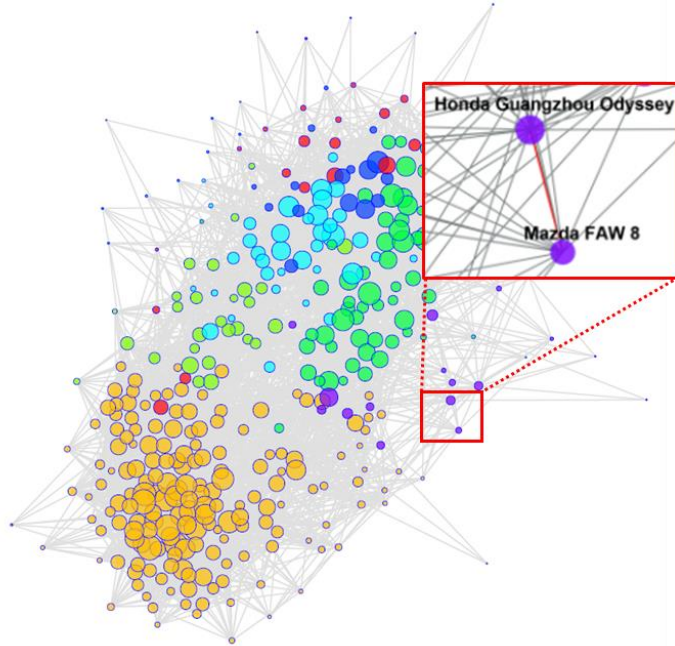
In the first implementation, we demonstrate the unidimensional network analysis (Sec. 3.2) for identifying aggregated product associations and hierarchical preference relations. Beyond existing literature, our work utilizes both consideration and purchase data in market surveys to derive relationships among vehicle products for understanding customer preferences and product competitions. We develop two types of product association networks – a *vehicle association network* with *undirected* links showing the similarity of products, and a *vehicle hierarchal network* with *directed* links indicating preference hierarchies.

The two vehicle association networks are constructed using the 2013 New Car Buyers Survey (NCBS) data provided by an independent research institute in China. The dataset contains 49,921 new car buyers who considered and purchased from a pool of 389 vehicle models in 2013. Both the set of considered vehicles and the final purchase are recorded for each customer. Customer demographics and product information are also reported by respondents.

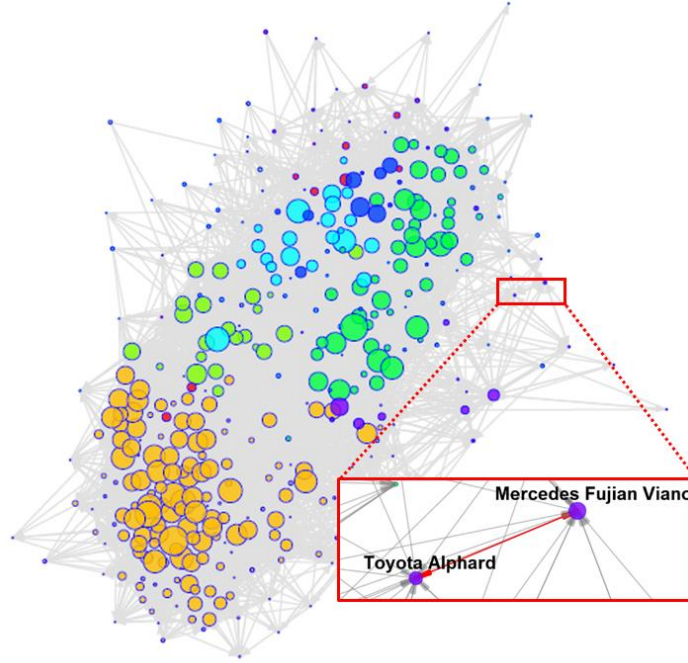
The vehicle association network is created to aid the analysis of customer consideration decisions by linking any pair of vehicles if both vehicles are considered by most consumers in his (her) consideration set. The association link is viewed as a form of similarity or closeness between any two vehicles in customers' minds. The link strength is quantified by *lift* to reflect how often the two products are compared by a population of customers. The lift between product i and product j is defined as the probability of co-consideration over the probability that they are being considered individually, see Eqn (2). The probability value is approximated by the percentage of product (co)occurrence recorded in the NCBS data.

$$lift(i, j) = \frac{\Pr\{\text{co-consider } i \text{ and } j\}}{\Pr\{\text{consider } i\} \cdot \Pr\{\text{consider } j\}}. \quad (2)$$

To prune the network links, a thinning threshold at 1 is chosen for the lift value; a lift greater than 1 has a precise statistical meaning showing a positive association between the two products (Tan et al., 2004). For example, Honda Guangzhou Odyssey and Mazda FAW 8 are positively associated, as shown in Fig. 7(a). The association link implies that the two products have a high chance of being co-considered. From the customer's perspective, it means that a customer considers Odyssey is also very likely to consider Mazda 8 at the same time.



(a). Centrality and Community in Vehicle Association Network.



(b). In-degree Hierarchy in Hierarchical Preference Network.

Figure 7. Unimodal Vehicle Networks Constructed from NCBS 2013. Nodes are Sized by Network Degrees (or In-degrees) and Colored by Network Communities. Network Layout is Computed by the Fruchterman–Reingold Force-Directed Algorithm based on Aesthetic Criteria (Fruchterman and Reingold, 1991). Link Strength is not Specified and has no relation to the Distance of Nodes.

As a measure of network *centrality*, the node degree calculates the number of links attached to a node. In the vehicle association network, products with a higher degree centrality are those frequently co-considered with many other vehicles by customers. Examples of high-degree centrality vehicles include GM SGM Chevrolet Sail, Audi FAW Q5, and Kia Dongfeng Yueda K2. One interesting observation is that most of the high-centrality vehicles are also among the most popular vehicles considered by customers, though the two quantities are not equivalent in definition. Another observation is that the node degrees is not uniformly distributed such that some vehicles are considered more frequently than others.

For the constructed vehicle network, the product *community* analysis is employed following Newman’s modularity method to determine groups of interconnected vehicles. In Fig. 7(a), the seven identified communities are marked in different colors. The product communities inform designer the marketing coverage of a brand family and marketing competence across several brands. For example, the yellow community includes most domestic entry-level sedans (e.g., BYD F6, Chery QQ, etc.), while the green community is featured by premium SUVs by foreign manufacturers (e.g., Jeep Grand Cherokee, Land Rover Discovery, etc.). It is also observed that a product line’s marketing success is highly influenced by its product positioning strategy. The successful product lines in the market generally cover more network communities. For example, as two marketing leaders in China, Volkswagen and GM

have covered 6 out of the 7 network communities, implying a great diversity of their vehicle products across multiple segments.

As a refinement to the above undirected network, a *directed network* is constructed where a link direction is determined through both consideration and purchase data in NCBS. If for any pair of vehicles, a customer considers both vehicles but chooses one over the other, the link direction will point towards the purchased vehicle. The lift metric shown in Eqn. (3) is slightly modified to accommodate the evaluation of directed link strength.

$$lift(i \rightarrow j) = \frac{\Pr\{\text{co-consider } i \& j, \text{purchase } j\}}{\Pr\{\text{consider } i\} \Pr\{\text{consider } j\}} \quad (3)$$

Again, the links are trimmed to highlight positive associations, where the link direction captures the preference hierarchy between the two linked products. For example, a bi-directional (mutual) link between Toyota Alphard and Mercedes Fujian Viano can be interpreted as the intense competition between the two products (Fig. 7(b)), because both vehicle models attract significant percentages of customers in considerations. Nevertheless, Viano gains a slight upper hand in market competition, because the strength of the link in that direction is stronger.

With a directed network, graph metrics indicating node *hierarchy*, such as node in-degree, can be computed to reflect customers' aggregated preferences across the population. The *in-degree* of a node computes the number of incoming links pointed to that node. A node with a high in-degree value implies the corresponding vehicle is very likely to be considered with other vehicles and is also more preferred in customer choice (purchase) decisions. For example, Audi FAW Q5 and Ford Kuga are popular vehicles in choice, which are ranked high in both degree centrality and in-degree hierarchy. In contrast, Volvo V40 and Ford Edge have been frequently considered (high degree centrality in undirected network), but fall behind in customers' final choices (low in-degree hierarchy).

Our illustrative example shows that descriptive network analysis may serve as a useful tool for designers to determine product positioning and product priorities in the phase of design planning. *Centrality*, *community*, and *hierarchy* allow designers to uncover the root-causes of the differences in vehicle sales under a specific market. These efforts may reveal issues that a design team could work on, e.g., product recognition (low centrality rank), coverage and diversity of product lines (products not appearing in certain communities), product competence (several vehicles in the same community), and product configuration (low hierarchy rank), etc.

While analyzing the structural information of a unidimensional network can be useful in describing product associations, there is a need for an approach to quantitatively evaluate customer heterogeneous preferences while addressing issues such as dependent alternatives, multiple decisions, social influence, and correlated observations. To demonstrate such capabilities of a network model, our next example employs ERGM in the MCPN framework with various nodes, relations and attributes included.

4.2. Using MCPN for Modeling Luxury Vehicle Preferences in Central China

Our second implementation demonstrates the use of inferential network technique (ERGM) for analyzing the vehicle MCPN framework (Secs 3.3 & 3.4). This network implementation also draws from the 2013 NCBS data to understand customer preference trends in China. With a focus on the luxury vehicle market, we examine respondents who live in the central provinces of China and consider only luxury imported vehicle models in their decision journey. This focused interest results in a subset data of 378 customers and 65 luxury vehicle models for modeling and

evaluation. As reported by McKinsey, the top reasons for Chinese customers to choose a luxury vehicle are: “reflection of social status”, “self-indulgence” and “business credibility”. Therefore, we expect that socially influenced decisions are more common in luxury vehicle buyers in China. In addition, the Chinese auto market is renowned for its complexity and volatility. Strong regional differences exist as a result for brand accessibility and lifestyle needs. Because of these hidden reasons beyond the functionality and design of a vehicle itself, quantifying the attractiveness of a vehicle attribute in such conditions becomes even more difficult.

The proposed MCPN integrates a feature-driven product association network, a customer-product network, and a customer social network as a unified entity for analysis. The implementation of the proposed approach goes beyond the descriptive analysis and consists of three major steps: network construction, ERGM specification, and ERGM interpretation; each of these steps is explained in the remaining of this section.

4.2.1. Data Transformation & Network Construction

1) Product Associations. Depending on the product complexity and the purpose of analysis, product associations can be built using either the “complete set of features” or “subsets”. In this example, product association links in the product layer of the multidimensional network are constructed using the complete set of attributes considered, including vehicle price, engine capacity, fuel consumption, and the existence of turbo. The association link is viewed as a form of overall product similarity from the perspective of engineering design. By converting the similarity of vehicle attributes as product association links, our emphasis in ERGM analysis is on testing whether customers tend to consider two vehicles with many common design features at the same time. Within the association network construct, the Gower's coefficient (Gower, 1971) is calculated to determine the existence of a link between any product pair. Gower's coefficient has the capability to appropriately handle continuous, ordinal, nominal and binary variables as inputs. Each continuous attribute is standardized by dividing each entry over the range of the corresponding attribute, after subtracting the minimum value; as a result, the Gower's similarity score has a range of [0, 1] exactly. Based on the empirical results, a global thinning threshold at 0.05 is chosen, which means the connected vehicles have similar levels across all attributes considered. This threshold also gives a reasonably dense network that ensures the estimated ERGM estimates are reliable (Lusher et al., 2012).

2) Preference Relations. We use the between-layer links connecting a product and a customer to model customers' consideration decisions over vehicle models. The structure of these links is precisely defined by NCBS data. In the survey, respondents are asked to report a list of vehicles that they seriously considered, including the purchased one. The number of consideration number ranges from 1 to 3. No customer listed more than 3 vehicles, even though the actual number might be higher.

3) Social Relations. Unlike the product association links which can be flexibly determined, the social links between customers have more specific meanings in social theories. Using the same strategy in our previous work on network simulation (He et al., 2014), a social space is constructed based on customer geographical locations and selected social attributes (age, income, education). Based on the homophily assumption that two customers with shorter distance in the social space are more likely to be connected, a global threshold is chosen to determine if a social link exists or not. To mimic the properties of real world networks, we then adjust the social links using the small-world model (Delre et al., 2007, Watts and Strogatz, 1998) to assure the high transitivity (“one's friends are

likely to be friends”) and low average path length (“six degrees of separation” between any two individuals). The small world mechanism provides a viable way to represent social links through both close and distant connections, implying that customers are not only influenced by their nearest neighbors in their social space but also a small number of remote contacts outside their regular social proximity.

Integrating the three types of network relations together, a visualization of the construction process for the MCPN structure is presented in Fig. 8. The complexity of network progressively increases from product association only in Fig. 8(a), to adding customer-product relations in Fig. 8(b), to adding the customer-customer relations in Fig. 8(c). As noted, we only include one type of preference link (consideration) and one type of product association link (feature-driven) for demonstration. All links are binary-valued and undirected.

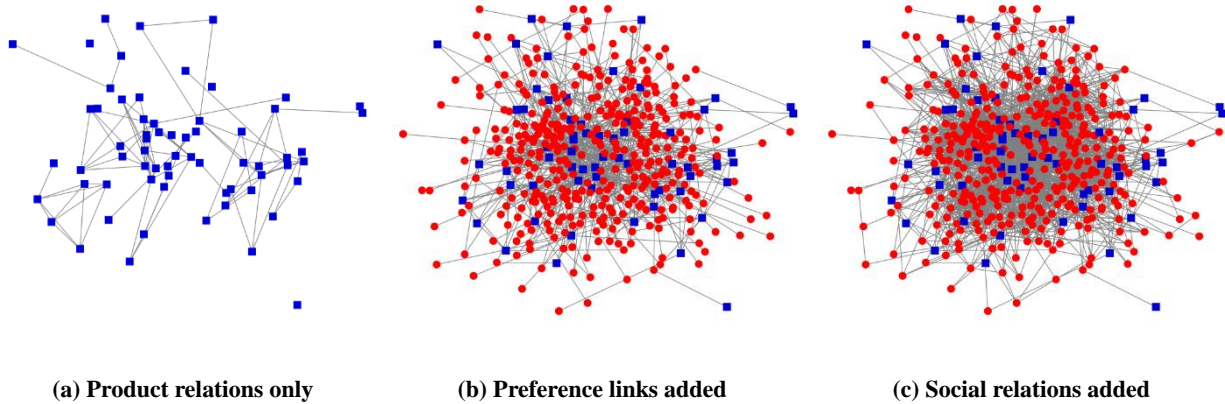


Figure 8: Progressive Construction of MCPN using NCBS 2013 Data. Products as Blue Squares and Customers as Red Disks.

4.2.2. Specification of ERGMs for Multidimensional Networks

With the constructed MCPN structure, the conditional form of ERGM is employed to address the question of how one or more dimensions of networks would affect the structures of other networks. Specifically, our research focus is on demonstrating the relevance and the feasibility of the network modeling technique. As presented in Table 4, the examined network effects are restricted to a subset of cross-level configurations and product/customer attributes of different forms. The choice of which network effect to include depends on the social theory, hypothesis, and the specific research questions to answer. Nevertheless, the demonstrated example serves as guidance for possible effects to consider in vehicle preference modeling for vehicle design.

4.2.3. Comparisons and Interpretations of ERGMs





Estimating the model coefficients for ERGM network effects is equivalent to fitting a model that gives maximal support to the data. However, the maximum likelihood estimates cannot be derived analytically due to the intractable constant in Eqn. (1) for a reasonable number of nodes. Thus, we employ a stochastic approximation (Snijders, 2002) that relies on MCMC simulations of graphs.




The result of the ERGM estimates for various specifications are presented in Table 4. The significant coefficient estimates are shown in bold font, meaning that the corresponding configurations are significant at the 95% confidence interval. We compare three model specifications based on the same dataset to highlight the benefits of the ERGM

approach. Model 1 formulates a bipartite ERGM analogous to a logistic model that contains only the attribute-relation effects composed by attributes of customers and products. This model allows the testing of influencing customer/product attributes in customer preference decisions, assuming that endogenous pure structural effects do not exist. Model 2 parameterizes a bipartite ERGM similar to Model 1 but with the addition of the pure structural effects and the cross-level product association effect. By comparing Model 2 and Model 1, one can test whether the addition of the pure structural effects and product association effect modify some of the attribute-relation effects in explaining customer preferences. The specification of Model 3 is the most complete model that includes all three types of ERGM effects. With the integration of the cross-level social influence effect, peer influences on preference decisions can be evaluated together with other product attributes, customer demographics, and structural patterns within the same model. Two penalized-likelihood criteria – AIC and BIC – are also provided in Table 4. The two measures decrease gradually as the addition of the considered structural effects, suggesting improved fits from Model 1 to Model 3.

The interpretation of Model 1 is similar to that for a logistic model. The vehicle *price* has a negative significant sign, implies that lower price is preferred in consideration of luxury vehicles. The significant positive *turbocharger* and *engine capacity* indicate that the presence of the turbocharger and the increased size of the engine would increase the probability for a customer to consider a particular vehicle model. The statistically negative *first-time buyer* suggests that first-time buyers are unlikely to enter the luxury vehicle market even though three out of four new cars are purchased by first-time buyers in China. The *fuel consumption* has a significant positive coefficient, meaning that fuel economy is less important for customers who decide to purchase a luxury vehicle. Interestingly, the decision of how many luxury vehicles to consider is less relevant to the household income, as seen by the insignificant *income* in the table. As noted, most model coefficients in Model 1 agree with our prior understanding about China's luxury market. This means that including attribute-related effects alone can capture an essential component of the process underlying the MCPN structure.

Table 4: Comparison of Three Specifications of ERGMs. For Each Considered Network Effect, the Graphical Configuration $z(y)$ is Presented Accompanied by the Estimated Coefficient (θ) and the Standard Error.

| | | Model 1 | | Model 2 | | Model 3 | |
|---|--|------------|------------|------------|------------|------------|------------|
| Configurations | Interpreted Effects | Est. Coeff | (Std. Err) | Est. Coeff | (Std. Err) | Est. Coeff | (Std. Err) |
| Pure Structure Effect | | | | | | | |
|  | Density | -7.0314 | (0.398) | -9.1009 | (0.495) | -8.9648 | (0.477) |
|  | Product popularity | | | 6.4955 | (0.644) | 6.5123 | (0.631) |
|  | Consideration range | | | -1.4036 | (0.516) | -1.3199 | (0.522) |
| Attribute-Relation Main Effect | | | | | | | |
|  | Price paid to the dealer (in 100K RMB) | -0.0346 | (0.020) | -0.0194 | (0.019) | -0.0182 | (0.018) |
| | Turbocharger (dummy) | 1.2796 | (0.109) | 1.0617 | (0.122) | 0.9056 | (0.118) |

| | | | | | | | |
|---|--|---------|---------|---------|---------|---------|---------|
| | Engine capacity (in cc) | 0.2809 | (0.134) | 0.2356 | (0.129) | 0.1871 | (0.119) |
| | Fuel consumption (in L/100km) | 0.1581 | (0.039) | 0.1270 | (0.036) | 0.1162 | (0.035) |
|  | First-time buyer (dummy) | -0.2343 | (0.096) | -0.9745 | (0.215) | -0.9744 | (0.214) |
| | Monthly household income (in 100K RMB) | 0.0027 | (0.002) | 0.0102 | (0.003) | 0.0125 | (0.003) |
| Cross - Level Effect | | | | | | | |
|  | Customer considers similar products | | | 0.9930 | (0.209) | 0.9704 | (0.212) |
|  | Peer influence | | | | | 0.4524 | (0.076) |
| Model Fit | | | | | | | |
| Null Deviance | | 34061 | | 34061 | | 34061 | |
| Residual Deviance | | 4847 | | 4831 | | 4773 | |
| AIC | | 5148 | | 4851 | | 4795 | |
| BIC | | 5205 | | 4932 | | 4884 | |
| Bolded coefficients are different from null at the 95% confidence interval. | | | | | | | |

In Model 2, the addition of the pure structural effects and the cross-level *customer considers similar products* effect considerably changes the interpretation of the underlying preference data. The significant positive *product popularity* indicates a dispersed degree distribution of product nodes. This implies customers' consideration decisions mostly concentrate on only a few vehicle models in the market. In contrast, the degree distribution is more centered for customer nodes, as shown by the negative *consideration range* coefficient, because customers only consider a limited number of vehicles (1-3) in NCBS data. The *customer considers similar products* effect is an indicator of how likely a customer may co-consider two vehicles that share similar design attributes. The significant positive coefficient means most people would judge a vehicle by its engineering attributes and consider multiple vehicles with similar levels of performances and prices. Concerning the attribute-relation effects, all the product effects (*turbocharger*, *engine capacity*, *fuel consumption*) generate smaller coefficients in magnitudes to their counterparts in Model 1 and the *price* is no longer significant. The change of *price* coefficient implies that price is actually not a decisive factor to consider for luxury vehicle buyers. In contrast, the customer effects of *first-time buyer* and *income* become more obvious. This is partly because the number of decisions (degree of customer nodes) has been controlled by the *consideration range*.

The coefficients of Model 3 are largely consistent with those in Model 2, except that the previously insignificant *income* becomes significant, while the previously significant *engine capacity* becomes insignificant. The significant positive *peer influence* indicates that a customer is likely to become “irrational” in decision making and simply considers what his/her peer has considered. Modeling the peer influence is a unique contribution of our work as such effect cannot be modeled either theoretically or computationally without the MCPN framework.

By comparing the above three models, several interesting findings can be summarized about the preference modeling in a multidimensional network context. First, including the attribute-relation main effects alone (Model 1)

can explain a large part of the formation of preference links. This observation is consistent to the foundational theory of many attribute-based preference modeling approach, such as DCA. Second, a model with only attribute-relation effects but no other relevant structural effects may ignore some of the underlying social structures represented by the structural patterns; therefore, such a model may produce biased results even if a researcher is only interested in a subset of product/customer attributes. For example, a popular product may attract a larger or smaller percentages of customers than we expected. This is captured by the *product popularity* (alternating k-stars) effect as a measure of node degree dispersion in a network model, but not possible in classical regression models. Finally, the peer influence effect (Model 3) introduces another layer of dependencies between two customers into the structure of the network. The significant positive estimate reflects the importance of social influence in explaining customer behavior and modeling product demand. Overall, the results of this example suggest that the nodal attributes (representing customer and product attributes) and network structures (representing product associations, social influences, and other underlying effects) are indispensable elements playing together in shaping the preferences of customers.

5. DISCUSSION AND CONCLUSION

While Discrete Choice Analysis (DCA) has been widely used to predict the influence of design decisions on customer preference and firm profit, in this paper, we introduce a conceptual framework of a drastically different approach using multidimensional network analysis (MNA) for modeling customer preferences in supporting engineering design decisions. We demonstrated the progression of a simple unidimensional network that contains only product associations, to a multidimensional network that considers product associations together with customer preference decisions, and finally to a more complete multidimensional structure that integrates product associations, customer social influence, and preference decisions as one network entity.

The descriptive network analysis as presented in the unidimensional network example offers a convenient tool to summarize key facts about the customer preference data. Through descriptive network measures, nodes can be clustered into subsets (community) or organized in ranks (centrality, hierarchy) to reflect structural positions in a network. When complex product association relationships are converted into market segments and competitive rankings, designers can better monitor product positions within a brand or between brand competitors. Next, the inferential network analysis with ERGM as illustrated in the MCPN framework enables the detailed modeling of both the network structures and customer/product attributes in a rigorous statistical sense. Compared to traditional logit models, the ERGM for MNA approach can handle complex relational data whose properties cannot be reduced to only the attributes. This capability resolves many issues in traditional preference models, as summarized as follows:

- Product associations can be modeled explicitly. In ERGM, product alternatives are no longer mutually exclusive, but interdependent in a network structure to influence customer’s preference decisions.
- Evaluation of social influence is enabled. By constructing customer social links in the customer layer, ERGM allows the social network effect to be statistically assessed and compared with other factors within a single model.
- Nested decisions can be analyzed through structural modeling. The model estimates can uncover not only a customer’s taste for a particular product, but also the relationship between several preference decisions as well as the number of decisions made, as represented by the correlated structural effects.

- Correlated product/customer attributes can be evaluated. Since ERGM assumes the observed network as a single realization from a multivariate distribution, no independence assumptions are necessary over the explanatory variables. Correlated product/customer attributes can be treated as structural terms and evaluated simultaneously.
- Coefficient estimates are highly interpretable and the ERGM results can be easily integrated into an engineering design optimization problem. The model estimates in ERGM resemble closely the outputs of DCA, enabling the assessment of various product configurations and their impacts on customer preferences.

The network-based preference model is superior in reducing design uncertainties, because it takes into account both customer and product attributes at the disaggregated level, and integrates customer decisions with product associations and social influences. In addition, the proposed modeling framework provides plentiful opportunities in engineering design research. The results of the China's luxury vehicle study have direct impacts on understanding customer consideration decisions over heterogeneous products in vehicle design. It can help automakers produce more competitive products in shorter times to market, considering not only the engineering requirements but also the heterogeneous preferences as well as the underlying social impacts. Although our approach is demonstrated for vehicle designs, the same principles and framework can be extended to other product designs and infrastructures involving consideration of many alternative options and heterogeneous customer preferences, such as designing electronic devices, software, transportation systems, energy supplies, etc. Designers can easily assess customers' willingness to choose innovative products over traditional products under social influence, and evaluating the benefits of introducing new designs to the market.

Though powerful and flexible, MNA has certain characteristics that need careful attentions when implemented for preference modeling in the context of product design. Depending on the purpose of the analysis, the size of a network model can vary from a few nodes to hundreds or thousands of nodes containing a diverse set of products. However, the network model could be sensitive to the issue of missing data and influenced by how links are defined (Kossinets and Watts, 2006). In addition, for a poorly specified model, degeneracy may occur in model estimation and cause the Markov chain to move towards an extreme graph of all or no edges (Snijders, 2002). This issue can be solved by incorporating curved exponential family terms that exhibit more stable behavior in model construction (Snijders et al., 2006).

As this paper is focused on developing the conceptual framework of the proposed approach, our next step is to enrich the case study by introducing more complex structural effects. The current MCPN application will be extended to incorporate other types of relations, e.g., directed association links for products and purchase decision links between customers and products. Examples of research questions to be answered may include how product associations and social relations may impact customers' preference decisions, and how customer preference decisions will in turn affect market competitions implied by product associations. In addition, we will devote more efforts to the tasks of network model evaluation and prediction. The use of network analysis for prediction is a new topic in the network research community. We will extend the multidimensional network model presented in this paper to predict customers' consideration sets and product choices. Changes of customer preference decisions will be forecasted under new design scenarios and market settings in order to translate the developed ERGM into methods suitable for design decision support.

ACKNOWLEDGMENTS

We thank Ford Corporate Economics and Global Consumer Insight team for providing the China market data. This work is financially supported by National Science Foundation (CMMI-1436658), Army Research Laboratory (W911NF-09-2-0053), and the Ford Northwestern Research Alliance program.

REFERENCES

- ARAL, S. & WALKER, D. 2011. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science*, 57, 1623-1639.
- BEN-AKIVA, M. E. & LERMAN, S. R. 1985. *Discrete choice analysis: theory and application to travel demand*, MIT press.
- CHEN, W., HOYLE, C. & WASSENAAR, H. J. 2013. *Decision-based design: Integrating consumer preferences into engineering design*, Springer.
- CHOWDHURY, G. 2010. *Introduction to modern information retrieval*, Facet publishing.
- CLAUSET, A., NEWMAN, M. E. J. & MOORE, C. 2004. Finding community structure in very large networks. *Physical review E*, 70, 066111.
- CONTRACTOR, N., MONGE, P. & LEONARDI, P. M. 2011. Network Theory| Multidimensional Networks and the Dynamics of Sociomateriality: Bringing Technology Inside the Network. *International Journal of Communication*, 5, 39.
- CORMIER, P., DEVENDORF, E. & LEWIS, K. Optimal Process Architectures for Distributed Design Using a Social Network Model. ASME 2012 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 2012. American Society of Mechanical Engineers, 485-495.
- COROMINAS-MURTRA, B., GONÍ, J., SOLÉ, R. V. & RODRÍGUEZ-CASO, C. 2013. On the origins of hierarchy in complex networks. *Proceedings of the National Academy of Sciences*, 110, 13316-13321.
- CRANMER, S. J. & DESMARAIS, B. A. 2011. Inferential network analysis with exponential random graph models. *Political Analysis*, 19, 66-86.
- DE VRIES, H. 1998. Finding a dominance order most consistent with a linear hierarchy: a new procedure and review. *Animal Behaviour*, 55, 827-843.
- DELRE, S. A., JAGER, W. & JANSSEN, M. A. 2007. Diffusion dynamics in small-world networks with heterogeneous consumers. *Computational and Mathematical Organization Theory*, 13, 185-202.
- DIMAGGIO, P., HARGITAI, E., NEUMAN, W. R. & ROBINSON, J. P. 2001. Social implications of the Internet. *Annual review of sociology*, 307-336.
- FAN, B., QI, G., HU, X. & YU, T. 2013. A network methodology for structure-oriented modular product platform planning. *Journal of Intelligent Manufacturing*, 1-18.
- FRANK, O. & STRAUSS, D. 1986. Markov graphs. *Journal of the American Statistical Association*, 81, 832-842.
- FREEMAN, L. C. 1979. Centrality in social networks conceptual clarification. *Social networks*, 1, 215-239.
- FRISCHKNECHT, B. D., WHITEFOOT, K. & PAPALAMBROS, P. Y. 2010. On the suitability of econometric demand models in design for market systems. *Journal of Mechanical Design*, 132, 121007.
- FRUCHTERMAN, T. M. J. & REINGOLD, E. M. 1991. Graph drawing by force - directed placement. *Software: Practice and experience*, 21, 1129-1164.
- GEYER, C. J. & THOMPSON, E. A. 1992. Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 657-699.
- GOWER, J. C. 1971. A general coefficient of similarity and some of its properties. *Biometrics*, 857-871.
- GREGORY, S. 2007. An algorithm to find overlapping community structure in networks. *Knowledge discovery in databases: PKDD 2007*. Springer.
- GUPTE, M., SHANKAR, P., LI, J., MUTHUKRISHNAN, S. & IFTODE, L. Finding hierarchy in directed online social networks. Proceedings of the 20th international conference on World wide web, 2011. ACM, 557-566.
- HE, L., CHEN, W., HOYLE, C. & YANNOU, B. 2012. Choice modeling for usage context-based design. *Journal of Mechanical Design*, 134, 031007.
- HE, L., WANG, M., CHEN, W. & CONZELMANN, G. 2014. Incorporating social impact on new product adoption in choice modeling: A case study in green vehicles. *Transportation Research Part D: Transport and Environment*, 32, 421-434.
- HOYLE, C., CHEN, W., WANG, N. & KOPPELMAN, F. S. 2010. Integrated Bayesian hierarchical choice modeling to capture heterogeneous consumer preferences in engineering design. *Journal of Mechanical Design*, 132, 121010.
- KARAHANNA, E., STRAUB, D. W. & CHERVANY, N. L. 1999. Information technology adoption across time: a cross-sectional comparison of pre-adoption and post-adoption beliefs. *MIS quarterly*, 183-213.
- KOSSINET, G. & WATTS, D. J. 2006. Empirical analysis of an evolving social network. *Science*, 311, 88-90.
- KRAUT, R. E., RICE, R. E., COOL, C. & FISH, R. S. 1998. Varieties of social influence: The role of utility and norms in the success of a new communication medium. *Organization Science*, 9, 437-453.

- LUSHER, D., KOSKINEN, J. & ROBINS, G. 2012. *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*, Cambridge University Press.
- MCPHERSON, M., SMITH-LOVIN, L. & COOK, J. M. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 415-444.
- MONGE, P. R. & CONTRACTOR, N. S. 2003. *Theories of communication networks*, Oxford University Press.
- MORROW, W. R., LONG, M. & MACDONALD, E. F. 2014. Market-System Design Optimization With Consider-Then-Choose Models. *Journal of Mechanical Design*, 136, 031003.
- NARAYAN, V., RAO, V. R. & SAUNDERS, C. 2011. How peer influence affects attribute preferences: A Bayesian updating mechanism. *Marketing Science*, 30, 368-384.
- NETZER, O., FELDMAN, R., GOLDENBERG, J. & FRESKO, M. 2012. Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31, 521-543.
- NEWMAN, M. E. & GIRVAN, M. 2004. Finding and evaluating community structure in networks. *Physical review E*, 69, 026113.
- PAGE, L., BRIN, S., MOTWANI, R. & WINOGRAD, T. 1999. The PageRank citation ranking: Bringing order to the web.
- PALLA, G., DERÉNYI, I., FARKAS, I. & VICSEK, T. 2005. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435, 814-818.
- REAL, R. & VARGAS, J. M. 1996. The probabilistic basis of Jaccard's index of similarity. *Systematic biology*, 380-385.
- RICE, R. E. & AYDIN, C. 1991. Attitudes toward new organizational technology: Network proximity as a mechanism for social information processing. *Administrative science quarterly*, 219-244.
- ROBINS, G., PATTISON, P., KALISH, Y. & LUSHER, D. 2007. An introduction to exponential random graph (p*) models for social networks. *Social networks*, 29, 173-191.
- SHIAU, C.-S. N. & MICHALEK, J. J. 2009. Should designers worry about market systems? *Journal of Mechanical Design*, 131, 011011.
- SNIJDERS, T. A. 2002. Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3, 1-40.
- SNIJDERS, T. A., PATTISON, P. E., ROBINS, G. L. & HANDCOCK, M. S. 2006. New specifications for exponential random graph models. *Sociological methodology*, 36, 99-153.
- SOSA, M., MIHM, J. & BROWNING, T. 2011. Degree distribution and quality in complex engineered systems. *Journal of Mechanical Design*, 133, 101008.
- SOSA, M. E., EPPINGER, S. D. & ROWLES, C. M. 2007. A network approach to define modularity of components in complex products. *Journal of mechanical design*, 129, 1118-1129.
- TAN, P.-N., KUMAR, V. & SRIVASTAVA, J. 2004. Selecting the right objective measure for association analysis. *Information Systems*, 29, 293-313.
- TRAIN, K. E. 2009. *Discrete choice methods with simulation*, Cambridge university press.
- TRIST, E. 1981. The evolution of socio-technical systems. *Occasional paper*, 2, 1981.
- ULRICH, K. T. 2003. *Product design and development*, Tata McGraw-Hill Education.
- URBERG, K. A. 1992. Locus of peer influence: Social crowd and best friend. *Journal of Youth and Adolescence*, 21, 439-450.
- WANG, M. & CHEN, W. 2015. A Data-Driven Network Analysis Approach to Predicting Customer Choice Sets for Choice Modeling in Engineering Design. *Journal of Mechanical Design*, 137, 071410.
- WANG, M., CHEN, W., FU, Y. & YANG, Y. 2015. Analyzing and Predicting Heterogeneous Customer Preferences in China's Auto Market Using Choice Modeling and Network Analysis. *SAE International Journal of Materials and Manufacturing*, 8.
- WANG, P., ROBINS, G., PATTISON, P. & LAZEGA, E. 2013. Exponential random graph models for multilevel networks. *Social Networks*, 35, 96-115.
- WASSERMAN, S. & FAUST, K. 1994. *Social network analysis: Methods and applications*, Cambridge university press.
- WATTS, D. J. & STROGATZ, S. H. 1998. Collective dynamics of 'small-world' networks. *nature*, 393, 440-442.