A Network-Based Approach to Modeling and Predicting Product Co-Consideration Relations

Zhenghui Sha
System Integration & Design Informatics Laboratory
University of Arkansas
Fayetteville, AR, USA

Yun Huang
Science of Networks in Communities
Northwestern University
Evanston, IL, USA

Jiawei Sophia Fu
Network for Nonprofit and Social Impact Laboratory
Northwestern University
Evanston, IL, USA

Mingxian Wang
Ford Analytics
Global Data Insight & Analytics
Ford Motor Company
Dearborn, MI, USA

Yan Fu
Ford Analytics
Global Data Insight & Analytics
Ford Motor Company
Dearborn, MI, USA

Noshir Contractor
Science of Networks in Communities
Northwestern University
Evanston, IL, USA

Wei Chen
Integrated Design Automation Laboratory
Northwestern University
Evanston, IL, USA

ABSTRACT

Customers often compare and evaluate alternative products before making final choices. Understanding customer preferences in consideration decisions is, therefore, an important step for choice modeling in engineering design. While the existing literature has shown that the exogenous effects (e.g., products' attributes and customers' demographics) are deciding factors in customers' consideration decisions, it is not clear how the endogenous effects (e.g., the inter-competitions among enterprises within a market) would influence customers' consideration decisions in a market. In order to address this issue, this paper presents a network-based approach using a statistical network model, called Exponential Random Graph Model, to study customers' consideration behaviors in supporting engineering design decisions. Our proposed network model is capable of characterizing the endogenous effects among products through various network structures (e.g., stars, triangles, etc.), and predicting whether customers would consider two products together (i.e., co-consideration), given the defined endogenous effects as well as the exogenous effects. In order to assess the performance of the proposed network model that considers endogenous effects, we compare it against the dyadic network model that only considers exogenous effects. Using the buyer survey data of China auto market in 2013 and 2014, we evaluate the goodness of fit and the predictive power of the two models at both the network level and the link level. The results show that our model has a better fit and predictive accuracy than the dyadic network model that only considers exogenous effects. This underscores the importance of the endogenous effects on customers' decisions at the consideration stage. The insights gained from this research help better understand the effects of market segmentation and product competition on customers' consideration decisions, and better explain how endogenous effects interact with the endogenous effects in affecting the customers' decision-making.

Keywords: Complex networks, Exponential Random Graph Model, consideration behavior, choice modeling, customer preference, engineering design.

1 INTRODUCTION

Complex network modeling and simulation have shown its power in many engineering applications, such as the wireless network, sensor network, smart grids, supply chain, transportation systems, and many others. Recent developments in mathematical modeling techniques and computational algorithms to study complex networks have drawn the attention of many fields, including engineering design. Complex networks have been used in engineering design for the study of relational patterns, effective network visualization of associations of products, and modeling social interactions [1] and cross-level interactions between customers and products [2, 3]. In the design of complex products, network analysis has been used to characterize a product as a network of components that share technical interfaces or connections. Various network metrics, such as clustering coefficients and path length, are used to characterize the product structure and study the correlations between design quality and the product structure. Based on the network metrics, e.g., the centrality, Sosa et al. [4] defined three measures of modularity as a way to improve the understanding of product architecture. Recent work by Sosa [5] 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 [6, 7] and improving multidisciplinary design efficiency [8]. In this paper, instead of focusing on the product or the designer, we leverage complex network modeling and simulation techniques to study another key stakeholder in product design – the customer. We aim to leverage complex networks to study customer preference in support of product design and development. Particularly, in this paper, we study customers' *consideration* decisions by modeling *product co-consideration relations* – two products are concurrently considered in purchase by customers – as a complex network.

2 BACKGROUND AND LITERATURE REVIEW

Choice modeling predicts product demand and market share as a function of engineering design attributes and customer profiles in a target market [9]. Choice models have been integrated into design optimization to take account of customer preferences in supporting engineering design decisions [9-12]. Previous choice models mostly assume that customers have bounded rationality and have underlying utilities to rank alternatives in a *consideration set* – "a set of product alternatives available to an individual who will seriously evaluate through comparisons before making a final choice" [13]. A key step of constructing choice models is to determine the consideration set [14]. As Hauser et al. [15] indicated, "if customers do not consider your product, they can't choose it."

From an enterprise perspective, understanding customer preferences in consideration is important for identifying crucial product features that customers are willing to pay for. Existing studies [16, 17] also revealed the *consideration set phenomenon*, i.e., the size of the consideration set tends to be much smaller (roughly 5-6 brands) than the total number of choices available in a market. As a result, small changes in individuals' consideration sets (either size or options) may significantly transform the landscape of the overall market and reshape the competition relations in an existing market. Therefore, understanding customers' preferences in consideration poses new opportunities to optimize product configurations, address customer needs, establish competitive design strategies, and make strategic moves such as branding and positioning.

Managerial actions have been taken to influence customers' consideration decisions directly, e.g., by changing brand accessibility [18] and by controlling usage and awareness [19]. However, quantitative studies on customers' consideration decisions are challenging as consideration is an intermediate construct, not the final choice [15]. The decision context and a large amount of uncertainty alter decision rules.

Existing literature primarily focuses on inferring decision-rule heuristics [20-22], such as the cognitive simplicity rule [23], which has been shown to be effective in automobile and Web-based purchasing. There are three approaches to uncover consideration decision-rule heuristics [15]. The first approach only utilizes final choices and product features in the consideration set. It adopts a two-stage consider-then-choose decision process and infers model parameters using the Bayesian or maximum likelihood estimation. Typical methods include Bayesian [24], choice-set explosion [25-27], and soft constraints [28]. The second approach measures consideration through designed experiments *in-vitro*, similar to the choice-based conjoint analysis exercise [15]. Then the decision rules that best explain the observed consideration decisions are estimated with Bayesian [29] and machine-learning pattern-matching algorithms [30]. The third approach measures decision rules directly through self-explicated questions [31].

Despite the diversity of research on consideration sets, few studies have focused on understanding the underlying process of generating customer consideration sets. The connection between the formation of consideration sets and the driving factors is not well understood. Particularly, we know little about how the inherent market structure, including both *the interdependence among existing products* and *association among customers,* affects the consideration decisions. To address this research gap, we develop a network-based approach to characterize customers' consideration behaviors through modeling product co-consideration relations. As shown in Figure 1, the key idea of the proposed network approach is to transform customer consideration sets into a product association network, in which nodes represent products and links represent the co-consideration between two products. As a result, the problem of understanding customer consideration can be addressed by predicting certain network structures as a function of association networks formed by product attributes and customer demographics. It is worth noting that as the link formation is an aggregation of customers' decisions, the links (i.e., the co-consideration relations) imply the competition relations among products. Therefore, our approach enables us to study customer preference and market structure in an integrated manner. This is different from the studies in choice modeling (e.g., the monomial logit choice model [32]) that focus on establishing models for individuals. It is also worth noting that our study is different from the agent-based models which hypothesize certain individual choice-making rules [33]. Instead, our approach is *data-driven*, which leverages the empirical observed data to drive the establishment of co-consideration models and prediction analysis using the estimated model parameters.

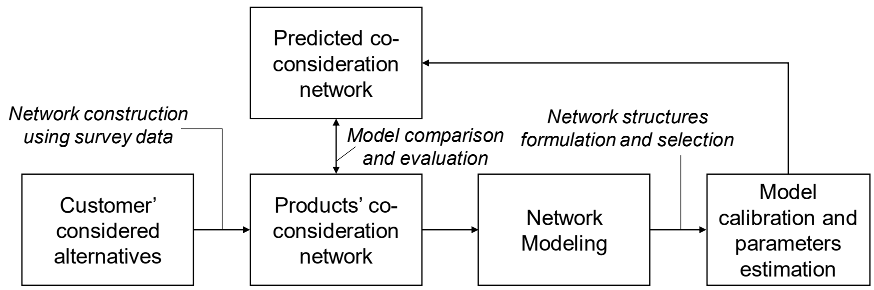


Figure 1. The research approach and the research focus

Network approaches have been also extensively used in recommender systems recently [34-38]. Recommender system is frequently used to recommend products to customers based on what they searched (considered). From the network representation point of view, our approach has the similarities to the bipartite projection approach [39] in the recommender system research. However, the proposed network approach is distinct from the network-based recommender algorithms [37, 38] mainly in two aspects: First, the end goal is different. The recommender algorithms attempt to **predict** future likes and interests by mining data on past user activities. Common methods include the similarity-based methods (e.g., the collaborative filtering [38], content-based analysis [40], Dirichlet allocation [41], etc.) and the recently developed hybrid methods [36, 42]. The approach proposed in this paper relies on the network-based *statistical inference model*, which emphasizes **deduction** and **explanation**. It aims to provide an explanatory framework for customers' consideration behaviors, so that a feedback loop can be created from customer preference to engineering design. Therefore, the end goal of this study is to inform product design for larger market share. In such a context, prediction is used in this study for comparison and validation purposes. Second, the role of network in the modeling is different. In existing network-based recommender algorithms, the input takes various graph-based node-specific attributes (e.g., degree), which are essentially the exogenous factors, to generate the similarity metrics. In our approach, the model input can take into account present network structures (e.g. triangles, loops), which represents the interdependencies among products, so that the effect of the inherent competition relations can be assessed. Such a capability supports better understanding on the consideration behaviors and could provide additional insights into the design research that has been primarily driven by users' preferences to engineering attributes.

The current work builds upon our previous research efforts. In our recent study, Fu et al. [43] developed a two-stage bipartite network modeling approach to study customer preference in making choices by decoupling the choice-making process in two stages, the consideration stage and the choice-making stage. To understand the underlying relations between product/customer attributes and customers' considerations, Wang et al. [44] utilized a *dyadic network* analysis approach (using multiple regression quadratic assignment procedure, MRQAP) to predict product co-consideration relations based on *exogenous factors,* such as product attributes and customer demographics. By mapping specific technological advancement (e.g., turbocharged techniques) to the change of products attributes, the authors also demonstrated how the model facilitates the forecast of the impact of technological changes on product co-consideration and market competition.

In this paper, we take a further step to investigate the power of complex network modeling in understanding product co-consideration relations by considering both *exogeous factors* and *endogenou factors*, e.g., product interdependence and inherent market competition. The core technique is based on the exponential random graph model (ERGM) [45]. While dyadic network models like MRQAP are convenient to predict the associations between products based on exogenous factors, ERGM incorporates endogenous factors and as well as other network interdepencies [46].

The **research objective** of this study is therefore two-fold: a) to establish the network-modeling framework that suppports the explanination of customer's consideration behaviors and enables the prediction of future market competitions; b) to compare the ERGM and dyadic network model to examine if the inclusion of product interdependence through the endogenous network effects would better capture the dynamics underlying the formation of product co-consideration relations. The remaining of the paper has five sections. Section 3 presents the research problem and introduces the method of constructing a product co-consideration network. We also briefly provide the technical background of the dyadic network model and ERGM. Section 4 describes the vehicle case study and the data source. We present the estimation results of the dyadic model and ERGM, and illustrate how to use the attribute-related network structures to represent product interdependence, i.e., the endogenous effects. To evaluate the performance of each model, Section 5 assesses model fit at both the global network level and at the local link level, and Section 6 evaluates the performance of each model in predicting future co-consideration relations. Finally, Section 7 presents the closing comments.

3 NETWORK CONSTRUCTION AND INTRODUCTION TO NETWORK MODELS

3.1 Network construction

The product co-consideration network is constructed using data from customers' consideration sets. The presence of a link (i.e., co-consideration) between two nodes (i.e., products) is determined by an association metric, called *lift* [47]. Equation (1) defines the *lift* value between products i and j. Similar to pointwise mutual information [48], *lift* measures the likelihood of the co-consideration of two products given their individual frequencies of considerations.

$$lift(i,j) = \frac{Pr(i,j)}{Pr(i) \cdot Pr(j)}, \tag{1}$$

where $Pr(i,j)$ is the probability of a pair of products i and j are co-considered by customers among all possibilities, calculated based on the collected consideration data; and $Pr(i)$ is the probability of individual product i being considered. The $lift$ value indicates how likely two products are co-considered by all customers at the aggregate level, normalized by the product popularity in the entire market. We use this probability of co-consideration, different from market share that is directly determined by the total purchases, to capture the competition between products. With the $lift$ value, an undirected co-consideration network can be constructed using the following binary rule:

$$E_{ij} = \begin{cases} 1, & if\ lift(i,j) \geq cutoff \\ 0, & otherwise \end{cases}, \tag{2}$$

where *cutoff* is the threshold to determine the presence of a link E_{ij} between two nodes i and j. Statistically, the $lift$ value 1 indicates that two products are completely independent [44]; a *lift* value greater than 1 indicates the two products are co-considered more likely than expected by chance. Based on the application context, research interest, and model requirement, different $lift$ values greater than 1 can be used as the cutoff value. Equations (1) and (2) suggest that the network adjacency matrix is symmetric and binary. In this paper, the research is focused on predicting whether two products would have been co-considered or not. The extent of how often they are co-considered (reflecting the competition intensity) is not the research focus of this paper. This is why we made the decision of using binary network instead of weighted network. Modeling a binary network, while computationally simpler, is not as rich as the valued network. Hence, we tested the robustness of our findings by estimating multiple models based on varying the cutoff values of *lift*, the intensity of co-consideration relations.

3.2 Research question in the network context

Once a co-consideration network is constructed, the likelihood of customers considering two products can be formulated as the probability of a co-consideration link. For prediction purpose, this leads to the question of what factors (e.g., product attributes and customer demographics) drive the formation of a link between a pair of nodes, and how significantly each factor plays a role in the link formation process. The aforementioned **research question** is recast as how to build a network model to predict whether a co-consideration link exists given the network structures, product attributes, and customer profiles.

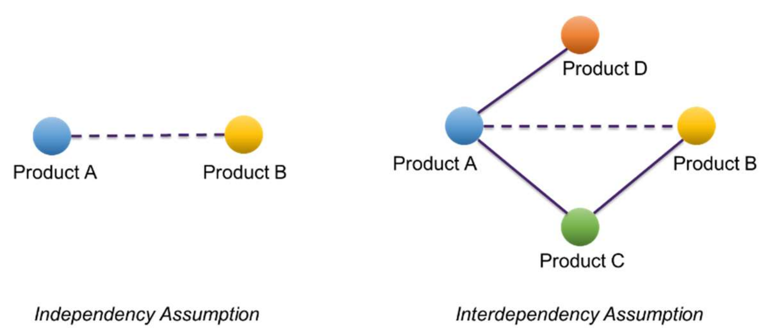


Figure 2. Two dependence assumptions underlying the co-consideration network

We posit that there are two decision-making scenarios underlying the co-consideration relations. The first scenario (Figure 2 on the left) assumes that each pair of products is independently evaluated by customers. Even for multiple alternatives in a consideration set, it treats the comparison of each two of these alternatives independent of other pairwise comparisons. The second scenario takes a more general interdependence assumption, where the formation of one co-consideration link is not independent of other co-consideration links. For example, in the right diagram of Figure 2, the likelihood of a co-consideration link between products A and B may be affected by the fact that they are both co-considered with product C. For the two aforementioned network models, the dyadic network model takes the simple independence assumption, while the ERGM assumes that all co-consideration relations sharing one node are interdependent. In this paper, we will examine whether the ERGM provides a more accurate understanding of the factors driving product co-considerations by evaluating the goodness of fit and the predicability of these two models.

3.3 Introduction to network models

The dyadic network model is analogous to the standard logistic regression element-wise on network matrices, where the model is given by:

$$logit\left[Pr\left(Y_{ij}=1\right)\right] = \boldsymbol{\beta}\boldsymbol{X}^{(\boldsymbol{n})} = \beta_0 + \beta_1 X_{ij}^{(1)} + \cdots + \beta_n X_{ij}^{(n)} \ . \tag{3}$$

The response Y_{ij} is the binary links E_{ij} between nodes i and j defined in Equation (2). The node attributes are converted to a vector of as *dyadic variable,* $\mathbf{X}^{(n)} = (x_{ij}^{(1)}, \dots, x_{ij}^{(n)})$. Each dyadic variable measures the similarity or difference between pairs of nodes based on the attributes of nodes and a specific arithmetic function (see Table 1 for various dyadic variables). The dyadic network models use the dyadic variables $\boldsymbol{X}$ to *predict* the complex structures of the observed network composed of co-consideration links. The coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_n)$ indicate the importance of individual dyadic variable in forming a co-consideration relation. Note that in this model, the probability of each link is evaluated independently.

3.3.1 Exponential random graph model

Other than the dyadic attribute effects, in a network, many links connected to the same node have endogenous relations. That means the emergence of a link is often related to other links. The ERGM introduced by [49, 50] is well known for its capability in modeling the interdependence among links in social networks. For example, two people who have a common friend are more likely to be friends with each other too, and therefore the three-person friendship relations form a triangle structure. Specific *network configurations*, including edges, stars, triangles, cycles, etc., can be used to represent different types of interdependence. The ERGM interprets the global network structure as a collective self-organized emergence of various local network configurations. The logic underlying ERGM is that it considers an observed network, $\boldsymbol{y}$, as one specific realization from a set of possible random networks, $\boldsymbol{Y}$, following the distribution in Equation (4) [45].

$$Pr(\boldsymbol{Y} = \boldsymbol{y}) = \frac{\exp\left(\boldsymbol{\theta}' \ \boldsymbol{g}(\boldsymbol{y})\right)}{\kappa(\boldsymbol{\theta}, \boldsymbol{y})}, \tag{4}$$

where $\boldsymbol{\theta}$ is a vector of model parameters, $\boldsymbol{g}(\boldsymbol{y})$ is a vector of the network statistics and attributes, and $\kappa(\boldsymbol{\theta}, \boldsymbol{y}) = \sum_{z \in y} \exp(\theta' g(z))$ is a normalizing quantity to ensure Equation (4) is a proper probability distribution. Equation (4) suggests that the probability of observing any particular network is proportional to the exponent of a weighted combination of network characteristics: one statistic $g(y)$ is more likely to occur if the corresponding θ is positive. Note that in ERGM, the network itself is a random variable and the probability is evaluated on the entire network instead of a link as it in Equation (3) for dyadic models. In brief, the advantages of using ERGM in the context of product co-consideration are three-fold: 1) using *network configurations* to characterize the endogenous effects among co-consideration links, 2) providing various dyadic variables to model different types of exogenous impacts of the product attributes, and 3) integrating both exogenous attribute effects and endogenous network effects in a unified framework.

3.3.2 Exogenous dyadic variables and endogenous network effects

The exogenous dyadic variables used both in the dyadic model and ERGM allow modeling of two types of effects between a pair of nodes with specific variables: the baseline effects of the attributes and the homophily effects, i.e., the similarity or difference between the attributes of two nodes [44, 51]. In the context of the product co-consideration network, the baseline effects examine whether products with a specific attribute are more likely to be co-considered than products without that attribute, e.g., imported car models could be more likely to be co-considered as compared to domestic car models. The homophily effects examine whether two products with similar attributes tend to have a co-consideration link. For example, customers are more likely to consider and compare products with similar prices. The development of dyadic variables supports the study of inherent product competition beyond the understanding of customer preferences.

Table 1 summarizes the guidelines of creating dyadic variables for different types of attributes such as binary, categorical, and continuous. For the product attributes under (a)-(c), the strength of the link X_{ij} is determined by the corresponding attributes x_i and x_j associated with the linked products. Beyond product attributes, we also introduce non-product related attributes (d). For example, customer demographics can be included in the model to allow the prediction of the impact of customers' associations/similarities on product co-consideration relations. To create a dyadic variables related to customers' attributes, multi-variable association techniques, e.g., joint correspondence analysis (JCA) [52], have been used to compute the similarity of the customer-related attributes as the distance between two product points (x_i and x_j) in a metric space. In this paper, we follow the method presented in [53] to develop two categories of distance variables – the distance of customer perceived characteristics and demographic distance. The customer perceived characteristics are user proposed tags to indicate their perceptions of the products, such as youthful, sophisticated, and business-oriented. Customer demographics include income and family information of the user groups of each car models. The inclusion of customer associations through these distance-based dyadic variables is a unique feature of our network modeling approach.

Table 1: Constructing explanatory dyadic attributes

Configuration	Statistic	Dyadic effects
(a) Binary product attributes		
Sum variable	$X_{ij} = x_i + x_j$	Attribute baseline effect
Matching variable	$X_{ij} = I\{x_i = x_j\}$	Homophily effect

(b) Categorical product attributes		
Matching variable	$X_{ij} = I\{x_i = x_j\}$	Homophily effect
(c) Continuous product attributes (standardized)		
Sum variable	$X_{ij} = x_i + x_j$	Attribute baseline effect
Difference variable	$X_{ij} = \lvert x_i - x_j \rvert$	Homophily effect
(d) Non- product related attributes		
Distance variable	$X_{ij} = \lVert x_i - x_j \rVert_2$	Homophily effect
• $I\{\cdot\}$ represents the indicator function. • $\lvert\cdot\rvert$ represents the absolute-value norm on the 1-dimension space. • $\lvert\lvert \cdot \rvert\rvert_2$ represents the L_2-norm on the n-dimension Euclidian space.		

Different from the dyadic models that can only consider exogenous dyadic effects, the ERGM supports the modeling of product interdependence regarding *endogenous network effects*. In this paper, we are particularly interested in two *network configurations*, the star-type interdependence and triangle-type interdependence [1]. The star structures (the left diagram in Figure 3) indicate that the probability of one focal product being co-considered with others is conditional on the number of existing co-consideration relations of that focal product (e.g., the node on the top in the figure has three co-consideration links). A positive star effect suggests that a product is more likely to be co-considered with another product if it is popular and already being co-considered with many others. The triangle structures (the right diagram in Figure 3) indicates that if two products are co-considered with the same set of other products, they are more likely to be mutually co-considered. Positive star effects could include stars with varying number of links (such as 2, 3, 4, 5 and perhaps many more). Likewise, a link could have many triangles by being linked with varying number of nodes (1, 2, 3, 4, 5, and perhaps many more). Both star and triangular effects imply multi-way product competition. To combine the effects of stars with multiple links and multiple triangles, we use two *network configurations* – the *geometrically weighted degrees* and the *geometrically weighted edgewise shared partner*, respectively [54].

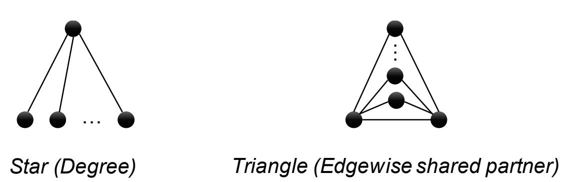


Figure 3. Two network configurations of co-consideration relations

4 CASE STUDY – MODELING VEHICLE CO-CONSIDERATION NETWORK

4.1 Application context and data source

When considering and purchasing a vehicle, customers make decisions on car models (e.g., Ford Fusion vs. Honda Accord), in part, based on their preferences for vehicle attributes (e.g., price, power, and make) and their demographics (e.g., income, age, etc.). To understand the effects of these factors on co-consideration relations among vehicle, we use data from a buyer survey in the 2013 China auto market. The dataset consists of about 50,000 new car buyers' responses to approximately 400 unique vehicle models. The survey covered a variety of questions, including respondent demographics, vehicle attributes, and customers' perceived vehicle characteristics. The respondents reported the car they purchased as well as the primary and secondary alternatives they considered before making the final purchase. These responses are used to construct the vehicle co-consideration network. The vehicle attributes reported in the survey are verified by vehicle catalog databases.

4.2 Vehicle co-consideration network

Following the method discussed in Section 3.1, we construct a vehicle co-consideration network with *cutoff* = 5 which results in a network of 389 nodes and 2,431 binary links. A smaller *cutoff* generated a denser network but had similar analytical results. We have tested our models using *cutoff* at 1, 3, 5, and 7 respectively, and no significant changes in the trends of the model results are observed. Figure 4 shows an example of a partial vehicle co-consideration network with 11 car models. The node size is proportional to the degree, and colors indicate the clusters in which the vehicles are more likely to be co-considered with each other. The number on each link is the *lift* value indicating the strength of the co-consideration.

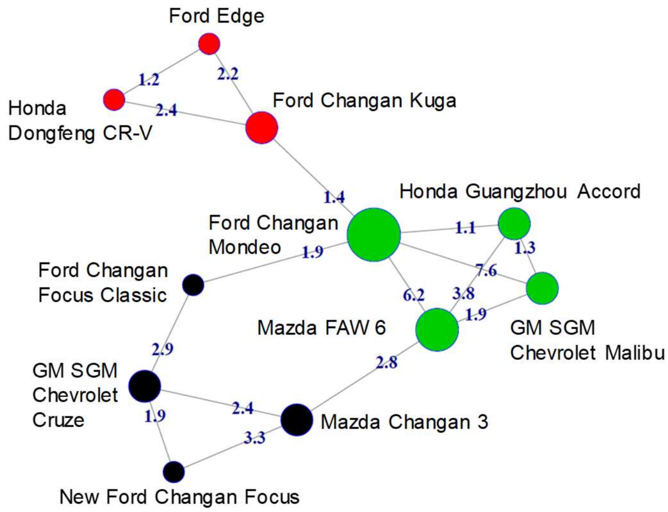


Figure 4. An example of partial vehicle co-consideration network

Table 2 summarizes some descriptive network characteristics. For example, the average degree suggests that on average each vehicle has 12.5 co-considered vehicles and indicates the overall intensity of competition in the market. The clustering coefficient (CC), on the other hand, measures the cohesion or segmentation of the vehicle market [44]. The average local CC at values of 0.26 indicates the strong cohesion embedded in the network, and vehicle models are frequently involved in multi-way co-consideration in the market. The descriptive network analysis facilitates the understanding of the auto market and provides guidelines on the selection of *network configurations* in ERGM.

Table 2: Representative network characteristics of the generated co-consideration network

Number of nodes	Number of links	Average degree	Average path length	Average local cluster coefficient
389	2,431	12.5	3.34	0.26

4.3 Descriptive statistics of the independent dyadic variables

Many exogenous dyadic variables related to vehicle attributes, such as the difference and sum variables of car prices, engine power, fuel consumption, and matching variables of vehical's market segments and make origin, could change the patterns of co-consideration among the vehicle models. We use infroamtion gain analysis to select 12 most important dyadic variables among all 22 possible dyadic variables. The log transformation (base 2) is applied to the price and engine power variables to offset the effect of large outliers. Table 3 shows the descriptive statistics of the independent variables.

Table 3. Descriptive statistics of independent variables for 389 car models in 2013

	Mean (SD)	Min	Max
Vehicle attributes			
Import (binary)	145 import & 244 domestic		
Price ($\log_2$)	17.61(1.34)	14.50	20.84
Power ($\log_2$)	7.27(0.58)	5.25	8.76
Fuel consumption (per 100 BHP)	6.61(1.62)	2.99	18.56
Market segment (categorical)	17 car segments		
Make origin (categorical)	13 American, 22 American-Chinese, 98 Chinese, 90 European, 50 European-Chinese, 31 Japanese, 54 Japanese-Chinese, 11 Korean, 20 Korean-Chinese		
Vehicle attribute matching and difference			
Market segment matching	10.1% pairs of cars co-considered are in the same segment		
Make origin matching	16.5% pairs of cars co-considered have the same make origin		
Price ($\log_2$) difference	1.53 (1.12)	0	6.34
Power ($\log_2$) difference	0.66 (0.49)	0	3.51
Fuel consumption difference	1.71 (1.52)	0	15.58
Customer association			
Distance of customers' perceived characteristics	0.20 (0.13)	0	1
Distance of customers' demographics	0.27 (0.16)	0	1

In total, six vehicle attributes, *import status*, *price*, *engine power*, *fuel consumption*, *market segment*, and vehicles' *make origin*, are considered in the model. *Import* is a binary variable describing whether a car is imported (*import* = 1, 37.3%) or domestically produced (*import* = 0, 62.7%). As suggested in Table 1 and Section 3.3.2, we construct a sum dyadic variable of *import* to account for its baseline effect of whether each paired cars are both imported (value 2 for 13.90% of the pairs), one imported and one domestic (value 1 for 46.76%), or both domestic (value 0 for 39.34%). If the baseline effect of the import attribute is positive, the coefficient of the sum variable of *import* should be positive as well, i.e., the higher the sum value of the two car models, the more likely they are co-considered together. Similarly, the sum variables of *price* (in RMB and transformed using $\log_2$) and *power* (in brake horsepower BHP and transformed using $\log_2$) describe the baseline effects of price and power on product co-consideration relations. We construct a variable, *fuel consumption*, by dividing litres of gasoline each vehicle consumed per 100 kilometers over vehicle power (in 100 BHP). As such, the smaller this value, the more fuel-efficient a car model is. The difference variables of price, power, and fuel consumption capture the homophily effects, which are used to test if the car models with similar attributes (smaller differences) are more likely to be co-considered together.

The auto industry is very competitive, so most car models have very clearly targeted customers and compete in a specific market segment. Since vehicle's *market segment* is a categorical variable, we use a dyadic matching variable in the model to investigate whether two cars from the same segment would affect their co-consideration patterns. The top 3 in all 17 segments in our sample are the C-Class Sedan (21.6% of car models), B-Class Sedan (11.3%), and Small Utility (11.1%). Similarly, *make origin* is also a categorical variable, and it describes the region where the car brand originates. Our dataset shows that 90, 31, 11, and 13 car models are made in Europe, Japan, South Korea, and the United States, respectively. While 98 car models are produced in China with local brands, other local-foreign joint venture brands come from Europe (50), Japan (54), South Korea (20), and the United States (22). The matching variables of *market segments* and *make origins* are used to account for people's homophily behavior of comparing cars with the same brand and origin.

4.4 Model implementation using ERGM

Table 4 shows the estimated coefficients and corresponding odds ratios from fitting the dyadic and ERGM models. Other than the variables described above, the ERGM includes three additional variables associated with *network configurations*. The *edge* variable controls the number of links to ensure the estimated networks have the same density as the observed one. Conceptually, if we have no knowledge about the cars' attributes or their co-consideration relations, the *edge* estimates the likelihood that two cars will be co-considered randomly, like an intercept term in a regression or a "base rate". The *star effect* and *triangle effect* discussed in Section 3.3.2 are measured by *geometrically weighted degree* and the *geometrically weighted edgewise shared partner*, respectively. According to the ERGM, most vehicle attributes, except the *price* baseline effect and *power* difference, are statistically significant (*p-value* < 0.001) and therefore play important roles in vehicle co-consideration. For instance, two vehicles with smaller differences in price and fuel consumption are more likely to be co-considered. If the price of one car model is twice the price of another car, their odds of co-consideration is only 45% of the odds of two cars with the same price. Similarly, one liter per 100km per 100 BHP difference in fuel consumption leads to 93% of the odds of co-consideration compared to the cars with the same fuel consumption. For the matching of vehicle attributes, two vehicles in the same market segment are 1.94 times more likely to be co-considered than the ones in different segments, and two vehicles with the same make origin are 1.69 times more likely to be co-considered than the ones with different origins. Finally, the negative coefficient for the distance of customers' demographics shows that customers with different demographics are less likely to co-consider the same vehicle. In summary, the results show that customers are more likely to consider cars with similar perceived features, such as price, fuel consumption, market segment and make origin.

As shown in Table 4, the coefficient of the *triangle effect* is 0.70 (*p-value* < 0.001). The positive sign indicates that two vehicles co-considered with the same set of vehicles are more likely to be co-considered with each other. It implies that a form of multi-way grouping and comparison exists in customers' consideration decisions, i.e. product alternatives in a person's consideration set are considered as the same time. On the other hand, the positive coefficient of the *star effect* (inversely measured by *geometrically weighted degree*) indicates that most of the cars tend to have a similar amount of co-consideration links. With these endogenous network effects, the ERGM significantly improves the model fit compared to the dyadic model as indicated by the improvement of BIC from 16,005 to 14,021. In the next section, we perform a systematic comparative analysis to evaluate how well the simulated networks match the observed vehicle co-consideration network.

Table 4: Estimated coefficients and odds ratios of the dyadic model and ERGM

	Dyadic Model		**ERGM**	
Input variables	**Est. coef.**	**Odds**	**Est. coef.**	**Odds**
Network configurations of product interdependence				
Edge / Intercept	-14.36**	0.00	-13.71**	0.00
Star effect (inverse measure)			1.97**	7.20
Triangle effect			0.70**	2.01
Baseline effects of vehicle attributes				
Import	0.37**	1.45	0.11**	1.11
Price ($\log_2$)	-0.02	0.98	-0.007	1.01
Power ($\log_2$)	0.68**	1.97	0.35**	1.42

Fuel consumption (per 100 BHP)	0.19**	1.21	0.12**	1.13
Homophily effects of vehicle attribute matching and difference				
Market segment matching	1.38**	3.98	0.66**	1.94
Make origin matching	1.28**	3.60	0.53**	1.69
Price difference ($\log_2$)	-1.75**	0.17	-0.80**	0.45
Power difference ($\log_2$)	0.08	1.09	0.13	1.14
Fuel consumption difference	-0.08*	0.92	-0.07**	0.93
Homophily effects of customer association				
Distance of customers' perceived characteristics.	-0.42	0.66	-0.31	0.74
Distance of customers' demographics	-0.57**	0.56	-0.37*	0.69
Model performance				
Null deviance	104,618			
Bayesian Information Criterion (BIC)	16,005		14,021	

Note: * *p-value < 0.01,* ** *p-value < 0.001*

5 MODEL COMPARISON ON GOODNESS OF FIT

A goodness of fit (GOF) analysis is performed to compare the model fit of dyadic and ERGM models. Using the dyadic and ERGM models in Equations (3) and (4), respectively, and based on the estimated parameters in Table 4, we compute the predicted probabilities of co-consideration between all pairs of vehicle models. The links with predicted probabilities higher than a threshold (e.g., 0.5) are considered as links that exist. Once the synthetic networks are obtained from both models, we compare them against the real 2013 co-consideration network at both the network level and the link level. The network level evaluation uses the *spectral goodness of fit* (SGOF) metric [55]; while the link level evaluation uses various accuracy measurements, such as *precision*, *recall*, and *F scores* (see Section 5.2 for more details).

5.1 Network-level comparison

Spectral goodness of fit (SGOF) is computed as:

$$SGOF = 1 - \frac{E\bar{S}D_{obs,fitted}}{E\bar{S}D_{obs,null}}, \tag{5}$$

where $E\bar{S}D_{obs,fitted}$ is the mean Euclidean spectral distance for the fitted model while $E\bar{S}D_{obs,null}$ is the mean Euclidean spectral distance for the null model, i.e., the Erdős–Rényi (ER) random network in which each link has a fixed probability of being present or absent. Hence, SGOF measures the amount of the observed structures explained by a fitted model, expressed as a percent improvement over a null model. The Euclidean spectral distance computes the L_2 norm (also called Euclidean norm) of the error between the observed network and all k simulated networks, i.e., $||\boldsymbol{\epsilon}_k||$, where error $\boldsymbol{\epsilon}$ is the absolute difference between the spectra of the observed network ($\hat{\boldsymbol{\lambda}}^{obs}$) and that of the simulated network ($\hat{\boldsymbol{\lambda}}^{sim}$), i.e., $\left|\hat{\boldsymbol{\lambda}}^{obs} - \hat{\boldsymbol{\lambda}}^{sim}\right|$. Since the calculation of the spectra $\hat{\boldsymbol{\lambda}}$ requires eigenvalues of the entire network's adjacent matrix, this evaluation is performed at the network level. When the fitted model exactly describes the data, SGOF reaches its maximum value 1. SGOF of zero means no improvement over the null model. The SGOF metric provides an overall comparison of different models. It is especially useful when a modeler is not clear about which network structural statistics are important in explaining the observed network. For example, in our co-consideration network, it is hard to tell which network metrics, such as the average path length or the average CC, are more important to the understanding of market structure. Under this circumstance, the SGOF provides a simple yet comprehensive evaluation. Table 5 lists the SGOF scores of both dyadic model and the ERGM. Based on 1,000 predicted networks from each model, the results of the mean, 5th, and 95th percentile of SGOF show that the ERGM significantly outperforms the dyadic model.

Table 5: Spectral goodness of fit results of the dyadic model and ERGM

	Dyadic model	**ERGM**
Mean SGOF (5th percentile, 95th percentile)	0.37 (0.31, 0.43)	0.63 (0.48, 0.76)

5.2 Link-level comparison

In addition to the network-level comparison, the predicted networks are also evaluated at the link level. We define a pair of vehicles with a co-consideration relation as *positive*, whereas the ones without links as *negative*. Therefore, the *true positive* (TP) is the number of links predicted as positive and also positive in the observed network; the *false positive* (FP) is the number of links predicted as positive but actually negative, i.e., wrong predictions of positives. Similarly, the *true negative* (TN) is the number of links predicted as negative and observed as negative; the *false negative* (FN) is the number of links predicted as negative but observed as positive. Taking 0.5 as the threshold of predicted probability (as it is used in the logistic function), we calculate the following three metrics to evaluate the performance of prediction for both dyadic model and ERGM. *Precision* is the fraction of true positive predictions among all positive predictions; *recall* is the fraction of true positive predictions over all positive observations; *F score* is the harmonic mean of *precision* and *recall* (see Table 6 for the formulas). These metrics are adopted because each of them reflects the capability of the model from different perspectives. It could be the case where the model predicts many links (e.g., all links are predicted in extreme cases and FP is high) so that the *precision* is low and the *recall* is high, while another model could preict very few links that leads to high FN, and thereforehigh *precision* and low *recall*. Therefore, using either *precision* or *recall* only partically reveals the model performance. Hence *F score* is often recommended as a fair measure because it considers both *precision* and *recall* and provides an average score. In this study, we use all three metrics together to provide a complete picture of the model performance.

Table 6: Results of various metrics for link-level comparison (predicted links based on threshold at 0.5)

Metrics	Dyadic model	ERGM
$Precision = TP/(TP + FP)$	0.594	0.543
$Recall = TP/(TP + FN)$	0.042	0.311
$F_\beta = \frac{(1+\beta^2) \times Precision \times Recall}{(\beta^2 \times Precision + Recall)}$	$F_{0.5} = 0.162$ $F_1 = 0.078$ $F_2 = 0.051$	$F_{0.5} = 0.473$ $F_1 = 0.396$ $F_2 = 0.340$

As shown in Table 6, almost all performance metrics suggest that ERGM outperforms the dyadic model. In particular, the *recall* of ERGM is significantly higher than that of the dyadic model. The dyadic model is only able to predict about 4.2% of co-consideration; whereas the *recall* of the ERGM can reach 31.1%. These results imply that the inclusion of product interdependence in ERGM indeed improves the model fit and better explains the observed product co-consideration relations. The only metric for which the dyadic model has a better value is the *precision*. At the threshold of probability is equal to 0.5, the dyadic model only predicted 170 links as positive in total, and 101 of them were correct. The small denominator in the *precision* formula, i.e., *TP*+*FP*=170*,* produces a larger *precision*.

Since different thresholds of the predicted probability will affect the value of *precision* and *recall*,In order to get a more comprehensive understanding, we evaluate the *precision-recall curve* [56] by altering the threshold from 0 to 1. The model that has a larger area under the curve (AUC) performs better [57]. When evaluating binary classifiers in an imbalanced dataset (with many more cases of one value for a variable than the other), which is the case here, Saito and Rehmsmeier [57] have demonstrated that the *precision-*

recall curve is more informative than other threshold curves, such as the receiver operating characteristic (ROC) curve[1]. Figure 5 shows that for any given recall value, the *precision* of ERGM is strictly higher than that of the dyadic model and the ERGM outperforms the dyadic model in the full spectrum of the threshold of probability.

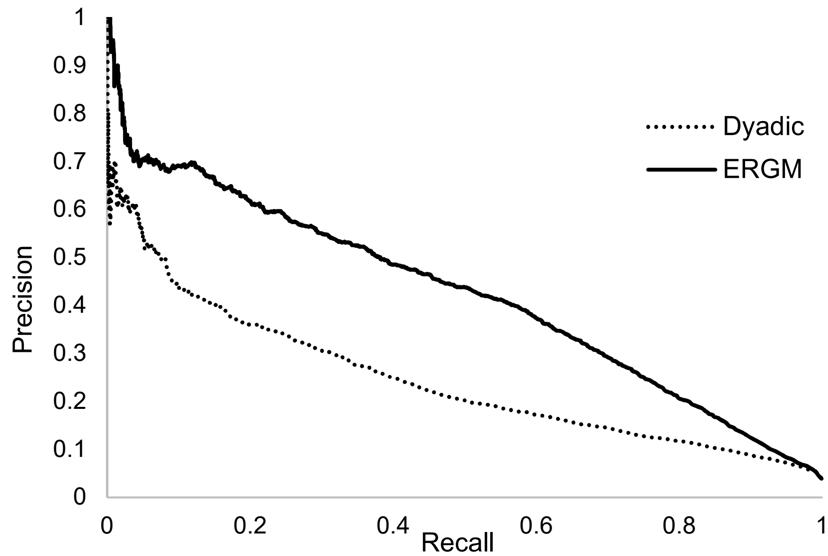


Figure 5. The precision-recall curve of the dyadic model and ERGM with random network benchmarked

In summary, the comparisons at both the network level and the link level validate our hypothesis that the product interdependence, i.e., the endogenous effect, plays a significant role in the formation of product co-consideration relations, and hence the customers' consideration decisions. In the next section, we examine the predicative power of the two models.

6 MODEL COMPARISON ON PREDICABILITY

In this section, we take a further step to compare the two models in terms of the predictability. We use the models developed with the 2013 dataset (i.e., the model coefficients shown in Table 4) to predict the vehicle co-consideration relations in the 2014 market. From an illustrative example in Figure 6, we can see some car models (e.g., node 4) withdrew from the market in 2014, some new car models (e.g., node 6 and node 7) were introduced to the market, but most of the car models (e.g., nodes 1, 2 3, and 5) remained in the

[1] We also studied the ROC curve, and drew the same conclusions

2014 market. In this paper, we focus on predicting the future co-considerations among the overlapping models in two consecutive years since the new models may introduce critical features not captured in the previous market, such as electric cars. In our study, 315 car models were available in both 2013 and 2014. Therefore, the task here is to predict whether each pair of cars among these 315 car models will be co-considered in 2014 given their new vehicle attributes in 2014, the new customer demographics, existing market competition structures[2], and the model coefficients estimated based on the 2013 data.

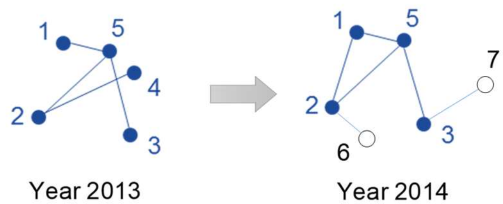


Figure 6. Illustration of the evolution of the co-consideration network

Most pairs of cars have the same dyadic status (i.e., co-considered or not) in 2013 and 2014. For example, if two car models were not co-considered in 2013, customers continued to not co-consider these two in 2014. This case is not of interest because predicting nonexistence is much easier due to the imbalance nature of the network dataset and it does not provide new insights. Similarly, the persistent co-consideration in both 2013 and 2014 is also expected. Therefore, we focus on changes in two prediction scenarios: emergence and disappearance of co-consideration links from 2013 to 2014. As shown in Table 7, among 47,724 pairs of cars that were not co-considered in 2013, 1,202 pairs were considered in 2014. The event of changing from not being co-considered to being co-considered indicates the change of market competition potentially caused by the change of vehicle attributes such as prices. On the other hand, 1,731 pairs of cars were co-considered in 2013 among the 315 car models, but 1,087 pairs were no longer co-considered in 2014. We indicate the two cases in the last column of Table 7 where the predictions of 2014 network using 2013 model are the events of interest. The two "Yes" cases, predicting emerging co-consideration and disappearing co-consideration links, both represent the change of co-consideration status

[2] The market competition structure is captured by the model coefficients of the three *network configurations* including the *edge*, *star effect* and *triangle effect* discussed in Section 4.4.

from 2013 to 2014 and are the positive outcomes of model predictions. Such predictions are more difficult (yet substantively more useful) to attain than the other two "No" cases of no-change. By testing both the dyadic and ERGM models, we examine which model had better predictive capability, assuming that the driving factors and customer preferences of co-consideration characterized by the model coefficients in Table 4 are unchanged from 2013 to 2014.

Table 7. Prediction scenarios of interest

Prediction scenarios	Year 2013	Year 2014	Events of interest
Emergence of co-consideration	47,724 pairs of cars not co-considered	1,202 pairs of new co-consideration	Yes
		46,522 no change	No
Disappearance of co-consideration	1,731 pairs of cars co-considered	1,087 pairs no longer co-considered	Yes
		644 no change	No

In both prediction scenarios, we input the new values of vehicle attributes and customer profile attributes from 2014 into the model. When using ERGM, characteristics of *network configurations* calculated based on the 2013 data also served as inputs for prediction. Once the models predict the probability of each pair of car models, we evaluate the performance metrics separately in two scenarios: 1) the *precision* and *recall* of predicting emerging co-consideration among the 47,724 pairs of not co-considered car models, and 2) the *precision* and *recall* of predicting the disappearance of co-consideration among 1,731 pairs of cars co-considered in 2013. The *precision* and *recall* of predictions are calculated similarly to the ones used in Section 5.2. The *precision* score is the ratio of the number of correctly predicted links (such as corrected prediction of emerging co-consideration or disappeared co-consideration) over the number of predictions a model makes. The *recall* score is the ratio of the number of correctly predicted links over the number of events of interest (true emerging co-consideration or disappeared co-consideration in 2014).

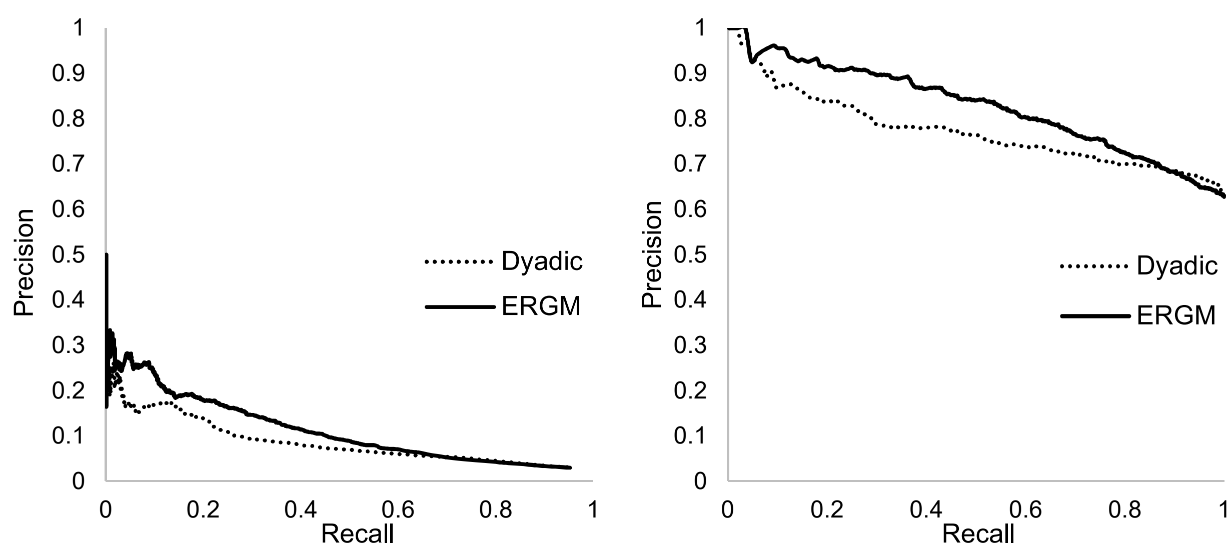
Table 8 shows the results of the prediction *precision* and *recall* calculated based on the predicted probability of 0.5 as the threshold in the two scenarios. To predict emerging co-consideration, the ERGM had much better performance than the dyadic model. Specifically, the dyadic model tends to be over-trained based on vehicle attributes and only predicts a small set of most likely links, i.e., 9 of the 1,202 emerging new co-

consideration relations. On the other hand, the ERGM predicted 111 (more than ten times) emerging co-consideration with the same *precision*. With the probability threshold of 0.5, the ERGM and dyadic model had similar differences in performance in predicting disappearing co-consideration links. The following Figure 7(b) shows that ERGM outperforms the dyadic model in almost all points of the *precision-recall* curve. In fact, the PR curves (Figure 7) show that ERGM at the entire range of the threshold outperforms the dyadic model in both prediction scenarios.

Table 8. The prediction precision and recall at the threshold of 0.5 in two prediction scenarios

Prediction scenarios	Model	# Events of interest (TP+FN)	# Predictions (TP+FP)	# Correct predictions (TP)	Prediction *precision*	Prediction *recall*	Prediction F_1
1	Dyadic	1202	36	9	0.250	0.0075	0.015
	ERGM		442	111	0.251	0.092	0.135
2	Dyadic	1087	1654	1076	0.651	0.990	0.785
	ERGM		1183	860	0.727	0.791	0.758

Therefore, we conclude that the ERGM has better predictability than the dyadic model. In addition to the GOF fitness test, the prediction test described above further validates our assumption that taking interdependencies in network modeling better explains the co-consideration network. In this particular case study, the analyses performed in both GOF and prediction tests indicate that vehicles' co-consideration relations are influenced by their existing competitions in the market.



(a) Predict emerging co-consideration *(b) Predict disappeared co-consideration*

Figure 7. Prediction PR curves of dyadic model and ERGM in two prediction scenarios

7 CLOSING COMMENTS

In this paper, we propose a network-based approach to study customers' preferences in consideration decisions. Specifically, we apply the *lift* association metric to convert customers' considerations into a product co-consideration network in which nodes present products and links represent co-consideration relations between products. With the created co-consideration networks, we adopt two network-modeling techniques, the dyadic model and the ERGM, to predict whether two products would have a co-consideration relation or not. Using vehicle design as a case study, we perform systematic studies to identify the significant factors influencing customers' co-consideration decisions. These factors include vehicle attributes (*price*, *power*, *fuel consumption*, *import status*, *make origin*, and *market segment*), the similarity of customer demographics, and existing competition structures (i.e., the interdependence among co-consideration choices captured by *network configurations*). Statistical regressions are performed to obtain the estimated parameters of both models, and comparative analyses are performed to evaluate the models' goodness of fit and predictive power in the context of vehicle co-consideration networks. Our results show that the ERGM outperforms the dyadic model in both GOF tests and the prediction analyses. This paper makes two contributions relevant to engineering design: a) a rigorous network-based analytical framework to study product co-consideration relations in support of engineering design decisions, and b) a systematic evaluation framework for comparing different network modeling techniques regarding GOF and prediction *precision* and *recall*.

This study provides three practical insights on co-consideration behavior in the China auto market. First, the customers are price-driven when considering potential car models. Both models suggest significant homophily effects of vehical prices and customer demographics in forming co-consideration links, i.e., car models with similar prices and targeting to similar demographics such as income and family size are more likely to be considered in the same consideration set. However, the ERGM reveals much more influencial drivers, such as the homophily effects of car segments and make origins. These findings comfirm the internal clusters in the auto market. Second, the ERGM model suggests that there were significantly fewer star strucutres but much more triangles in the co-consideration network. Beyond the impacts of the vehicle and customer attributes, ERGM also illustrates car models that received an equal amount of consideration and those that were likely to get involved in multi-way co-consideration. Third, the model comparisons with the goodness of fit analysis and prediction scenarios demonstrate that a network modeling approach that captures the interdependence of co-consideration, e.g., the ERGM approach, helps improve the predictability of product co-considerations.

Finally, having an analytical model in this application context could boost future explorations including the what-if scenario analysis that aims to forecast market responses under different settings of existing product attributes, as demonstrated in [44]. Since ERGM has a better model fit and predictability, it will help make more accurate projections on the future market trends and aid the prioritization of product features in satisfying customers' needs as well as supporting engineering design and product development. In the future, we plan to extend the network approach to a longitudinal weighted network-modeling framework, which not only predicts the existence of a link but also the strength of the co-consideration between car models in future years. The weighted network models would help discover the nuance in different customers' consideration sets and therefore provide more in-depth insights for product design and market forecasting.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support from NSF CMMI-1436658 and Ford-Northwestern Alliance Project. An early version of part of this work was presented in the 2017 International Conference on Engineering Design [58].

STATEMENT OF CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper. Funding sources mentioned in acknowledgement do not lead to any conflict of interests regarding the publication of this manuscript.

REFERENCES

1. Robins, G., et al., *Recent developments in exponential random graph (p*) models for social networks.* Social networks, 2007. **29**(2): p. 192-215.
2. Wang, M., et al. *A Multidimensional Network Approach for Modeling Customer-Product Relations in Engineering Design*. in *ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. 2015. American Society of Mechanical Engineers.
3. Wang, P., et al., *Exponential random graph models for multilevel networks.* Social Networks, 2013. **35**(1): p. 96-115.
4. Sosa, M.E., S.D. Eppinger, and C.M. Rowles, *A network approach to define modularity of components in complex products.* Journal of mechanical design, 2007. **129**: p. 1118.
5. Sosa, M., J. Mihm, and T. Browning, *Degree distribution and quality in complex engineered systems.* Journal of Mechanical Design, 2011. **133**: p. 101008.
6. Byler, E. *Cultivating the growth of complex systems using emergent behaviours of engineering processes*. in *International conference on complex systems: control and modeling*. 2000. Russian Academy of Sciences.
7. Contractor, N., P.R. Monge, and P. Leonardi, *Multidimensional networks and the dynamics of sociomateriality: Bringing technology inside the network.* International Journal of Communication, 2011. **5**(0).
8. Cormier, P., E. Devendorf, and K. Lewis. *Optimal Process Architectures for Distributed Design Using a Social Network Model*. in *ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. 2012. Chicago, IL: American Society of Mechanical Engineers.
9. Chen, W., C. Hoyle, and H.J. Wassenaar, *Decision-based Design : Integrating Consumer Preferences into Engineering Design*. 2013th ed. 2012: Springer.
10. Michalek, J.J., F.M. Feinberg, and P.Y. Papalambros, *Linking marketing and engineering product design decisions via analytical target cascading.* Journal of Product Innovation Management, 2005. **22**(1): p. 42-62.
11. Sha, Z., et al., *Modeling Airlines' Decisions on City-Pair Route Selection Using Discrete Choice Models.* Journal of Air Transportation, 2016. **24**(3): p. 63-73.
12. Sha, Z. and J.H. Panchal, *Estimating Local Decision-Making Behavior in Complex Evolutionary Systems.* Journal of Mechanical Design, 2014. **136**(6): p. 061003-061003.
13. Wang, M. and W. Chen, *A Data-Driven Network Analysis Approach to Predicting Customer Choice Sets for Choice Modeling in Engineering Design.* Journal of Mechanical Design, 2015. **137**(7): p. 071410-071410.
14. Sha, Z. and J.H. Panchal. *Estimating linking preferences and behaviors of autonomous systems in the Internet using a discrete choice model*. in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2014.
15. Hauser, J., M. Ding, and S.P. Gaskin. *Non-compensatory (and compensatory) models of consideration-set decisions*. in *2009 Sawtooth Software Conference Proceedings, Sequin WA*. 2009.
16. Shocker, A.D., et al., *Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions.* Marketing letters, 1991. **2**(3): p. 181-197.
17. Hauser, J.R. and B. Wernerfelt, *An evaluation cost model of consideration sets.* Journal of consumer research, 1990. **16**(4): p. 393-408.
18. Nedungadi, P., *Recall and consumer consideration sets: Influencing choice without altering brand evaluations.* Journal of consumer research, 1990. **17**(3): p. 263-276.
19. Srivastava, R.K., M.I. Alpert, and A.D. Shocker, *A customer-oriented approach for determining market structures.* The Journal of Marketing, 1984: p. 32-45.
20. Frederick, S., *Automated choice heuristics.* 2002.

21. Shao, W., *Consumer Decision-Making: An Empirical Exploration of Multi-Phased Decision Processes*. 2006, Griffith University Australia.
22. Yee, M., et al., *Greedoid-based noncompensatory inference.* Marketing Science, 2007. **26**(4): p. 532-549.
23. Hauser, J.R., et al., *Disjunctions of conjunctions, cognitive simplicity, and consideration sets.* Journal of Marketing Research, 2010. **47**(3): p. 485-496.
24. Gilbride, T.J. and G.M. Allenby, *Estimating heterogeneous EBA and economic screening rule choice models.* Marketing Science, 2006. **25**(5): p. 494-509.
25. Andrews, R.L. and T. Srinivasan, *Studying consideration effects in empirical choice models using scanner panel data.* Journal of Marketing Research, 1995: p. 30-41.
26. Chiang, J., S. Chib, and C. Narasimhan, *Markov chain Monte Carlo and models of consideration set and parameter heterogeneity.* Journal of Econometrics, 1998. **89**(1): p. 223-248.
27. Erdem, T. and J. Swait, *Brand credibility, brand consideration, and choice.* Journal of consumer research, 2004. **31**(1): p. 191-198.
28. Swait, J., *A non-compensatory choice model incorporating attribute cutoffs.* Transportation Research Part B: Methodological, 2001. **35**(10): p. 903-928.
29. Gilbride, T.J. and G.M. Allenby, *A choice model with conjunctive, disjunctive, and compensatory screening rules.* Marketing Science, 2004. **23**(3): p. 391-406.
30. Boros, E., et al., *An implementation of logical analysis of data.* IEEE Transactions on Knowledge and Data Engineering, 2000. **12**(2): p. 292-306.
31. Ding, M., *An incentive-aligned mechanism for conjoint analysis.* Journal of Marketing Research, 2007. **44**(2): p. 214-223.
32. Sha, Z., et al., *Analyzing Customer Preference to Product Optional Features in Supporting Product Configuration.* SAE International Journal of Materials and Manufacturing, 2017. **10**(2017-01-0243).
33. Eliaz, K. and R. Spiegler, *Consideration sets and competitive marketing.* The Review of Economic Studies, 2011. **78**(1): p. 235-262.
34. Resnick, P. and H.R. Varian, *Recommender systems.* Commun. ACM, 1997. **40**(3): p. 56-58.
35. Adomavicius, G. and A. Tuzhilin, *Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions.* Ieee Transactions on Knowledge and Data Engineering, 2005. **17**(6): p. 734-749.
36. Zhou, T., et al., *Solving the apparent diversity-accuracy dilemma of recommender systems.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(10): p. 4511-4515.
37. Yu, F., et al., *Network-based recommendation algorithms: A review.* Physica a-Statistical Mechanics and Its Applications, 2016. **452**: p. 192-208.
38. Lu, L.Y., et al., *Recommender systems.* Physics Reports-Review Section of Physics Letters, 2012. **519**(1): p. 1-49.
39. Zhou, T., et al., *Bipartite network projection and personal recommendation.* Physical Review E, 2007. **76**(4).
40. Pazzani, M.J. and D. Billsus, *Content-Based Recommendation Systems*, in *The Adaptive Web: Methods and Strategies of Web Personalization*, P. Brusilovsky, A. Kobsa, and W. Nejdl, Editors. 2007, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 325-341.
41. Blei, D.M., A.Y. Ng, and M.I. Jordan, *Latent Dirichlet allocation.* Journal of Machine Learning Research, 2003. **3**(4-5): p. 993-1022.
42. Fiasconaro, A., et al., *Hybrid recommendation methods in complex networks.* Physical Review E, 2015. **92**(1).
43. Fu, J.S., et al. *Modeling Customer Choice Preferences in Engineering Design Using Bipartite Network Analysis*. in *2017 ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*. 2017. Cleveland, OH: ASME.

44. Wang, M., et al. *Forecasting Technological Impacts on Customers' Co-consideration Behaviors: A Data-driven Network Analysis Approach*. in *42nd ASME Design Automation Conference*. 2016. Charlotte, NC: ASME.
45. Robins, G., et al., *An introduction to exponential random graph (p*) models for social networks.* Social networks, 2007. **29**(2): p. 173-191.
46. Shumate, M. and E.T. Palazzolo, *Exponential random graph (p*) models as a method for social network analysis in communication research.* Communication Methods and Measures, 2010. **4**(4): p. 341-371.
47. Tufféry, S., *Data mining and statistics for decision making*. Vol. 2. 2011: Wiley Chichester.
48. Church, K.W. and P. Hanks, *Word association norms, mutual information, and lexicography.* Computational linguistics, 1990. **16**(1): p. 22-29.
49. Frank, O. and D. Strauss, *Markov graphs.* Journal of the american Statistical association, 1986. **81**(395): p. 832-842.
50. Wasserman, S. and P. Pattison, *Logit models and logistic regressions for social networks: I. An introduction to Markov graphs andp.* Psychometrika, 1996. **61**(3): p. 401-425.
51. McPherson, M., L. Smith-Lovin, and J.M. Cook, *Birds of a feather: Homophily in social networks.* Annual review of sociology, 2001: p. 415-444.
52. Greenacre, M., *Correspondence analysis in practice*. 2017: CRC press.
53. Wang, M., et al. *A Network Approach for Understanding and Analyzing Product Co-Consideration Relations in Engineering Design*. in *DS 84: Proceedings of the DESIGN 2016 14th International Design Conference*. 2016.
54. Hunter, D.R., *Curved exponential family models for social networks.* Social networks, 2007. **29**(2): p. 216-230.
55. Shore, J. and B. Lubin, *Spectral goodness of fit for network models.* Social Networks, 2015. **43**: p. 16-27.
56. Powers, D.M., *Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.* 2011.
57. Saito, T. and M. Rehmsmeier, *The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets.* PloS one, 2015. **10**(3): p. e0118432.
58. Sha, Z., et al. *Modeling Product Co-Consideration Relations: A Comparative Study of Two Network Models*. in *the 21st International Conference on Engineering Design, ICED17*. 2017.