21ST INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED17
21-25 AUGUST 2017, THE UNIVERSITY OF BRITISH COLUMBIA, VANCOUVER, CANADA

MODELING PRODUCT CO-CONSIDERATION RELATIONS: A COMPARATIVE STUDY OF TWO NETWORK MODELS

Sha, Zhenghui (1); Wang, Mingxian (2); Huang, Yun (1); Contractor, Noshir (1); Fu, Yan (2); Chen, Wei (1)
1: Northwestern University, United States of America; 2: Ford Motor Company, United States of America

Abstract

Customers often compare and evaluate alternative products before making purchase decisions. Understanding customer preference is an important step for choice modeling in engineering design. This study presents a network approach to model co-consideration relations between products in supporting engineering design decisions. The network approach of co-consideration represents each product as a node, and a link between two nodes implies the two products are co-considered by customers. We compare two network-based modeling techniques – the multiple regression quadratic assignment procedure (MRQAP) and the exponential random graph model (ERGM). Using vehicle purchase data in the 2013 China market, we evaluate the goodness-of-fit of the two techniques at both network level and link level. The analysis indicates that the ERGM outperforms the MRQAP model. Specifically, the ERGM is able to characterize the interdependence of product co-considerations through various network configurations and therefore has a better fit of the data. The insights of co-consideration models help to understand market segmentation and product competitions as well as other types of product associations.

Keywords: Market implications, User centred design, Design methods, Customer preference, Complex networks

Contact:
Prof. Wei Chen
Northwestern University
Mechanical Engineering
United States of America
weichen@northwestern.edu

Please cite this paper as:
Surnames, Initials: *Title of paper*. In: Proceedings of the 21st International Conference on Engineering Design (ICED17), Vol. 6: Design Information and Knowledge, Vancouver, Canada, 21.-25.08.2017.

1 INTRODUCTION

Choice modeling considers the prediction of product demand and market share as a function of engineering design attributes and target market description (Chen et al., 2012). Design optimization utilizes choice modeling techniques to estimate customer preferences and support engineering design decisions (Chen et al., 2012; Michalek et al., 2005; Sha et al., 2016; Sha and Panchal, 2014b). Previous choice models mostly assume that customers have bounded rationality and have underlying utilities to rank alternatives in a consideration set. So a key step of constructing choice models is first to determine the consideration set (Sha and Panchal, 2014a), which has become a central topic in understanding customers' behavior (Carson and Louviere, 2014). By definition, customers' *consideration set*, also known as choice set (Ben-Akiva and Lerman, 1985) or evoked set (Howard and Sheth, 1969), is "a set of product alternatives available to an individual who will seriously evaluate through comparisons before making a final choice" (Wang and Chen, 2015). As (Hauser et al., 2009) 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 at the early stage of purchase. Existing studies (Shocker et al., 1991; Hauser and Wernerfelt, 1990) have 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 the market. As a result, small changes in individuals' consideration sets (either size or choices) may significantly transform the overall market landscape and reshape the competition relations in an existing market. Therefore, understanding customers' decision-making in consideration poses new opportunities to optimize product configurations, address customer needs, establish competitive design strategies, and make strategic enterprise moves such as branding and positioning.

Despite a variety of studies on consideration set, few studies focus on the underlying process of generating customer consideration sets. The connection between the formation of consideration sets and the driving factors associated with both customer and product attributes is not well understood. We know little about how the inherent market structure, including *both the interdependence among existing products and association among customers*, would affect the consideration decisions. To address this research gap, we develop a network-based approach to quantitatively understand customer's 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 becomes the prediction of certain network structures in such an association network using product attributes and customer demographics as well as similarity networks derived from these attributes. It is worth noting that our approach is different from the agent-based models in which individual choice behavioral rules are hypothesized, e.g., (Eliaz and Spiegler, 2011). Instead, our approach leverages the observed data to drive the establishment of co-consideration models and analysis.

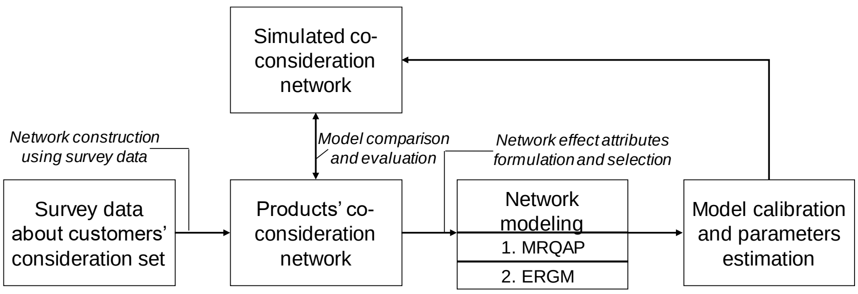

Figure 1. The research approach and research focus

Network analysis has many benefits, for example, in the study of relational patterns, effective network visualization of associations, and modeling social interactions (Robins et al., 2007b) and cross-level interactions (Wang et al., 2015; Wang et al., 2013). Our recent work (Wang et al., 2016b) developed a network-based analysis approach to forecasting the impact of technological changes on market competitions using the multiple regression quadratic assignment procedure (MRQAP). This study builds upon our previous efforts in modeling the underlying relations between product/customer attributes and customers' considerations. Specifically, we investigate the performance of a new modeling technique based on the exponential random graph model (ERGM) (Robins et al., 2007a), which takes both products interdependence and customers' associations into consideration. Although MRQAP is very convenient to test the association between networks, ERGMs better handle attribute variables, interdependent relations, and skewness in the distribution of network observations (Shumate and Palazzolo, 2010). The **research objective** is to compare the ERGM with the MRQAP model and to quantitatively evaluate how the inclusion of product interdependence would improve the model fit thus better predict the co-consideration.

The paper is structured as follows. Section 2 presents the research problem and introduces the method of constructing a product co-consideration network. We also briefly give the technical background of the MRQAP model and ERGM in this section. Section 3 describes the vehicle case study and the data source. Section 4 presents the model implementation and estimation results. We also present how the attribute-related network structures are configured and used to represent customers' associations and product interdependence. To evaluate the performance of each model, we use the estimated model parameters to regenerate co-consideration networks, compare the simulated networks with the real network, and assess the goodness-of-fit at both network level and link level. Finally, Section 5 presents the closing comments.

2 NETWORK CONSTRUCTION AND INTRODUCTION TO NETWORK MODELS

2.1 Network construction

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

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

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 market. Consideration probability is different from market share that is directly determined by the total purchases. With the $lift$ value, an undirected co-consideration network is constructed in a binary setting with the following rule:

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

where *cutoff* is a subjective threshold to determine the presence of a link E_{ij} between two nodes i and j. Statistically, a $lift$ value equals 1 indicates that two products are completely independent (Wang et al., 2016b); a *lift* value greater than 1 indicates the two vehicles are co-considered more likely than expected by chance, and vice versa. Based on the application context, research interest and model requirement, different $lift$ values greater than 1 can be selected. Equations (1) and (2) suggest that the network adjacency matrix is symmetric and binary.

2.2 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. That means we are interested in understanding what factors (e.g., the 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 **research question** in the network context is, therefore, how to predict whether a co-consideration link exists given the available data of an observed network, product profiles, and customer information.

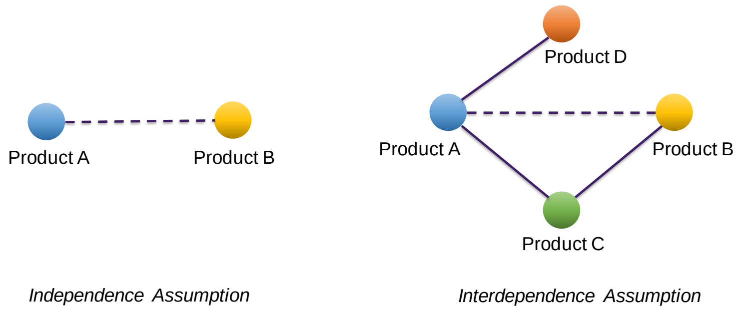

Figure 2. Two assumptions on the decision-making process underlying the co-consideration network

We posit two decision-making scenarios underlying the co-consideration relations. The first scenario (Figure 2 on the left) assumes that customers evaluate each pair of products independently. Even for multiple alternatives in a consideration set, it treats the comparison of each two of these alternatives independent of other pair-wise comparisons. The second scenario takes a more generic interdependence assumption, and therefore the formation of a co-consideration link may be due to the existing relations among the products. For example, in the right subfigure 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. The MRQAP model takes the simple independence assumption, while the ERGM assumes that all co-consideration relations sharing one node are interdependent. By evaluating the goodness of fit of two models, we examine whether the ERGM provides a more accurate understanding of the factors driving product consideration. The following subsection introduces the technical details of the two models.

2.3 Introduction to network models

2.3.1 Multiple regression quadratic assignment procedure (MRQAP)

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

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

The response Y_{ij} is the binary links E_{ij} between nodes i and j defined in Equation (2). The node (product) attributes are vectorized as *effect network* $\mathbf{X}^{(n)}=(x_{ij}^{(1)},\ldots,x_{ij}^{(n)})$, each measures the associations between pairs of nodes based on various arithmetic operations of attributes (see details in (Wang et al., 2016b)). The unique aspect of MRQAP is to use simple networks $\boldsymbol{X}$ (created using attribute data) to *predict* the structure of the observed complex decision network composed of co-consideration links. The coefficients $\boldsymbol{\beta}=(\beta_0,\beta_1,\ldots,\beta_n)$ in a MRQAP model indicate the importance of individual effect networks in forming a co-consideration relation. MRQAP permutes the rows and columns of the network matrix many times to generates a random null model, which is used to estimate the unbiased standard errors and pseudo p-values. Therefore, MRQAP is more accurate than the traditional regression model in the network context. Note that in this model, the probability of each link is evaluated independently.

2.3.2 Exponential random graph model

The ERGM was first introduced by (Frank and Strauss, 1986; Wasserman and Pattison, 1996) and is well known for its capability in modeling the interdependence among links in social networks. The emergence of a link in a network is often related to other links. For example, two people who have a common friend are very likely to be friends with each other too, and therefore the three friendship relations form a triangle structure. *Network configurations* capture local network structures, including edge, stars, triangles, cycles, etc., represent interdependence between network links. Then, the ERGM interprets the global network structure as a collective result of various local network configurations. The key logic behind 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) (Robins et al., 2007a).

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

where $\boldsymbol{\theta}$ is a vector of model parameters, $\boldsymbol{g}(\boldsymbol{y})$ is a vector of the network statistics, 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 $\boldsymbol{g}(\boldsymbol{y})$ is more likely to occur if the corresponding $\boldsymbol{\theta}$ 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. In brief, the advantages of using ERGM in the product co-consideration context have three aspects: 1) interdependence, 2) richness of explanatory variables, and 3) capability of characterizing both local and global network structures.

2.3.3 Effect networks and network configurations

The explanatory effect networks allow the modeling of two types of effects: the attribute-based main effect and the homophily effect (Mcpherson et al., 2001; Wang et al., 2016b). The attribute-based main effect tests whether products with a specific attribute is more likely to have consideration links than products without the attribute. The homophily effect represents the tendency of entities to associate and bond with similar others. In a product co-consideration network, the homophily effect tests whether two products with similar attributes tend to have a co-consideration link. A detailed description of the effect networks is presented in the Table 1 of (Wang et al., 2016b). The development of the effect network supports the study of embedded product competition beyond the understanding of customers' behaviors. In this paper, we also follow the method presented in (Wang et al., 2016a) to develop two distance networks – the customer demographic distance and perceived product characteristics distance – to capture the effect of customers' associations. The inclusion of customer associations through these distance networks is a unique feature of our approach.
Different from MRQAP that can only take effect networks, the ERGM also supports the modeling of product interdependence regarding network configurations. In this paper, we are particularly interested in two network configurations, the star-type interdependence and edgewise shared partner interdependence (Robins et al., 2007b), as shown in Figure 3. The alternating k-start indicates that the probability of two products A and B being co-considered is conditional on the number of existing co-consideration relations between product A and other products. The edgewise shared partner accounts for the dependent effect arising from the shared event. In co-consideration network, the inclusion of this effect helps answer the question: if two products are co-considered with the same product, are they more likely to have a second co-considered product in common, and a third one and so on?

Alternating k-star *Edgewise shared partner*

Figure 3. Two assumptions on the decision-making process underlying the co-consideration network

3 CASE STUDY – MODELING VEHICLE CO-CONSIDERATION NETWORK

3.1 Application context and data source

The application in this study is about vehicle consideration and purchase where customers make purchase decisions on a *car model* (e.g., Ford Fusion vs. Honda Accord), according to their preferences to various vehicle attributes (e.g., make, price and engine size) and their demographics (e.g., income, age, family size, etc.). The dataset used is a 2013 survey data on new car buyers in China auto market. The dataset consists of about 50k new car buyers' responses to about 400 unique vehicle models. The survey has questions covering a variety of topics, including respondent demographics, vehicle attributes, and customers' perceived vehicle characteristics (e.g., youthful, sophisticated, and business-oriented). The respondents were also asked to list the car they purchased, the main alternative car they considered, and the second considered alternative before making the final purchase. These responses are used to derive the co-consideration network being analyzed in this study. The vehicle attributes reported in the survey and verified by vehicle catalog database.

3.2 Vehicle co-consideration network

Following the method introduced in Section 2.1, we construct the vehicle co-consideration network with *cutoff* = 5. This value is determined based on the convergent performance of the ERGM. Smaller *cutoff* results in denser network (the network has 6449 links if *cutoff* = 1) that makes the computation of ERGM parameters hard to converge. With *cutoff* = 5, we obtained an unweighted network with 389 nodes and 2431 links. Table 1 lists some network metrics which may have different implications for the vehicle market. For example, the average degree measures the average number of co-considered vehicles each vehicle has and implies the average intensity of competition in the market. The clustering coefficient (CC) metrics, on the other hand, measures the cohesion or segmentation of the vehicle market (Wang et al., 2016b). The average local CC and the global CC at values of 0.26 and 0.28 indicate 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 our understanding of the market and provides guidelines on the selection of network configurations in ERGM.

Table 1. Representative network metrics of the generated co-consideration network

Number of nodes	Number of links	Average degree	Average path length	Average local cluster coefficient	Global cluster coefficient
389	2431	12.5	3.34	0.26	0.28

3.3 Model implementation

We create 25 effect networks, including difference and sum networks of price and fuel consumption, match networks of vehicle origin and turbo features, etc., and use information gain analysis to select 11 important effect networks to predict co-consideration links. The log transformation of some continuous variables is employed to offset the effect of large differences in certain car attributes. Besides, the ERGM includes three additional variables associated with network configurations. The *edge* variable is a network configuration controlling the number of links to ensure the estimated networks have the same density as the observed one.

Table 3 shows the estimated coefficients and corresponding odds ratios of the MRQAP and ERGM models. According to the ERGM, that most vehicle attributes, except the power difference and vehicle's import status, are statistically significant at the level of significance 0.01 and therefore have 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 of the price of another car, their odds of co-consideration is only 1.4% of the odds of two cars with the same price. Similarly, one mile per gallon difference in fuel consumption leads to 26.1% of the odds of co-consideration compared to the cars with the same fuel consumption. Similarly, for the matching of vehicle attributes, two vehicles in the same market segment are 1.804 times more likely to be co-considered than the ones in different segments, and two vehicles manufactured in the same county origin are 1.782 times more likely to be co-considered than the ones with different origins. The negative coefficient for the distance of customers' demographics shows that if the two vehicles targetted to customers with different demographics are less likely to be co-considered. In summary, the results show that customers are more likely to consider cars with similar features, such as price, fuel consumption, market segment, origin, and targeted demographics.

The results of the MRQAP model are consistent with the ERGM and have the exact sign and similar magnitude of the estimated parameters. However, the MRQAP has much bigger standard errors and tends to over-estimate the effects. The comparison of the MRQAP model and the ERGM shows that ERGM can take the product interdependence into account as network configurations while MRQAP does not have such a capability. As shown in Table 2, the coefficient of the shared partner distribution is 0.681, with p-values less than 0.001. This indicates that two vehicles co-considered with the same set of vehicles are more likely to be co-considered. It implies that a customer's consideration decision is also influenced by how the alternatives in his/her consideration set correlate (e.g., co-considered) with the products out of his/her entire consideration set. Such an observation is also evident in the performance improvement of ERGM (BIC=13997) as opposed to that of the MRQAP (BIC=15949).

Besides the BIC measure of model performance, in Section 4, we perform a systematic model comparison study and evaluate how well the observed vehicle co-consideration network would be recovered based on the two models.

Table 2. Estimated coefficients and odds ratios of the MRQAP model and ERGM

	MRQAP Model		**ERGM**	
Input variables	**Est. coef.**	**Odds**	**Est. coef.**	**Odds**
Intercept	-3.01*	0.05*		
Effect Networks of Vehicle Attributes				
Difference network of price w/ log transformation	-10.80*	2.05e-5*	-4.27*	1.4e-2*
Sum network of price w/ log transformation	1.67	5.32	0.60*	1.82*
Difference network of power w/ log transformation	1.30	3.67	0.56	1.76
Sum network of power w/ log transformation	-1.62	0.20	-0.88*	0.42*
Difference network of fuel consumption	-3.22*	0.04*	-1.34*	0.26*
Sum network of fuel consumption	2.77*	15.93*	1.32*	3.76*
Match network of a vehicle’s market segment	1.25*	3.50*	0.59*	1.80*
Match network of a vehicle’s origin	1.35*	3.85*	0.59*	1.78*
Match network of a vehicle’s import status	-0.07	0.93	-0.06	0.94
Effect Networks of Customer Association				
Distance network of customers’ perceived char.	-0.25	0.78	-0.19	0.83
Distance network of customers’ demographics	-0.59	0.55	-0.37*	0.69*
Network Configurations of Product Interdependence				
edge	N/A	N/A	-7.79*	0.4e-3*
Geometrically weighted degree	N/A	N/A	2.32*	10.15*
Geometrically weighted edgewise shared partner	N/A	N/A	0.68*	1.98*
Model performance				
Null deviance	104618			
Bayesian Information Criterion (BIC)	15949		13997	

* *indicates the estimated parameter is significantly different from 0 at the level of significance of 0. 01*

4 MODEL COMPARISON AND EVALUATION

To further compare the results of MRQAP and ERGM models, the goodness-of-fit (GOF) analysis is performed to test their predictive capability. Using the MRQAP and ERGM models in Equations (3) and (4) and estimated parameters in Table 2, we test whether the predicted probabilities of co-consideration between pairs of vehicles match with the real network constructed based on the 2013 NSCS data. The links with predicted probabilities higher than a threshold (e.g. 0.5) are accepted as predicted co-consideration relations between two vehicles. We compare the predicted network with the real network for each modeling technique at both the network level and the link level. The network level evaluation used the spectral goodness-of-fit (SGOF) metric (Shore and Lubin, 2015); the link level uses various accuracy measurements, such as precision, recall, and F scores (see Section 4.2 for more details).

4.1 Network-level comparison

The calculation SGOF follows Equation (5).

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

where $E\bar{S}D_{obs,fitted}$ is the mean Euclidean spectral distance under the fitted model while $E\bar{S}D_{obs,null}$ is the mean Euclidean spectral distance under the null model, i.e., the ER random model. Hence SGOF measures the amount of observed structure explained by a fitted model, expressed as a percent improvement over a null model. The calculation of Euclidean spectral distance takes the entire network adjacency matrix as an input, thereby the evaluation is performed at network level. SGOF is bounded above by 1, when the fitted model exactly describes the data. SGOF of zero means no improvement over the null model. The SGOF metric provides an overall comparison of different models, and one metric that 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 case, it is hard to tell which network metrics, such as the average path length or the average CC, are more important in understanding the market structure. Under this circumstance, the SGOF could be a less risky substitute. Table 3 lists the SGOF results of both MRQAP model and the ERGM. The results based on 1000 predicted networks from each of the two models include the mean, 5th, and 95th percentile of SGOF and show that the ERGM outperforms the MRQAP model, and it is statistically significant.

Table 3. Spectral goodness-of-fit results of the MRQAP model and ERGM

	MRQAP model	**ERGM**
Mean SGOF (5th percentile, 95th percentile)	0.35 (0.28, 0.42)	0.69 (0.60, 0.76)

4.2 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) means the number of links predicted as positive and observed as positive in the real network; the "false positive" (FP) means the number of links predicted as positive but observed as negative. Similarly, the "true negative" (TN) means the number of links predicted as negative and observed as negative; the "false negative" (FN) means the number of links predicted as negative but observed as positive. Taking 0.5 as the threshold of predicted probability, we calculate various metrics (shown in Table 4) to evaluate the performance of prediction for both MRQAP and ERGM.

Table 4. Results of various metrics for link-level comparison (predicted links based on threshold at 0.5)

Metrics	**MRQAP model**	**ERGM**
$Precision = TP/(TP + FP)$	0.641	0.546
$Recall = TP/(TP + FN)$	0.0448	0.316
$Accuracy = (TP + TN)/(TP + FN + FP + TN)$	0.968	0.969
$F_\beta = \dfrac{(1 + \beta^2)\times Precision\times Recall}{(\beta^2\times Precision + Recall)}$	$F_{0.5} = 0.21$; $F_1 = 0.084$; $F_2 = 0.055$	$F_{0.5} = 0.87$; $F_1 = 0.40$; $F_2 = 0.35$

Almost all performance metrics suggest that ERGM outperforms the MRQAP model. In particular, the recall of ERGM is significantly higher than that of the MRQAP model. The recall in our application is important because it measures the percentage of correctly predicted co-consideration relations among all 2431 observed co-consideration links. The MRQAP model is only able to predict about 4.5% of co-consideration; whereas the recall of the ERGM can reach 31.6%. These results imply that the inclusion of product interdependence in ERGM indeed improves the model fit and helps better explain the observed product co-consideration relations. The only metric of which the MRQAP has a larger value is the "precision." This is because at the threshold of probability equal to 0.5, the MRQAP only predicted 170 links as positive in total, out of which 109 links are correctly predicted. The small denominator of the precision formula *(TP+FP)* tends to produce a bigger precision. So, different thresholds of the predicted probability will affect the value of precision and recall. To get a comprehensive understanding of the recall and precision performance, we plot the precision-recall curve (Powers, 2011). The model that has a larger area under the curve performs better (Saito and Rehmsmeier, 2015). Figure 4 clearly shows that while both network models perform much better than the ER random network model, the ERGM overall outperforms the MRQAP model in the full spectrum of the threshold of probability.

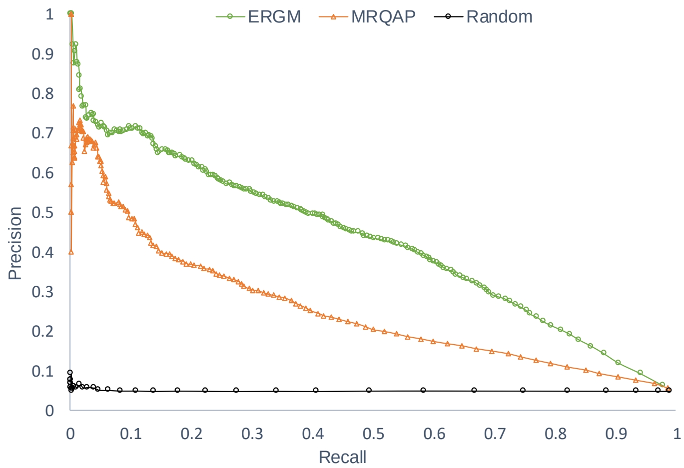

Figure 4. The precision-recall curve of the MRQAP model and ERGM with random network benchmarked

In summary, the comparative study performed at both the network level and the link level validates our hypothesis that the product interdependence plays a significant role in product co-consideration relations, and hence the customers' consideration behavior. The proposed ERGM technique is capable of modeling such interdependence and quantitatively capture the importance of various interdependence settings in forming a co-consideration relation between two products. Obtaining an analytical model in this application context could boost many future explorations including the what-if scenario analysis that aims to forecast market responses under different scenarios of existing product attributes, as what we have demonstrated in (Wang et al., 2016b). A better model enhances the predictive capability and is expected to make a more accurate projection of the future market trends, and aid the prioritization of product features in satisfying customers' needs and supporting engineering design and product development.

5 CLOSING COMMENTS

The proposed network approach and the evaluation methods provide a rigorous analytical framework to study the customer's co-consideration decisions. The approach uses network as an abstraction of system structures, thus is domain-independent. For example, it can be applied to non-engineering fields, such as social networks, to study the effect of associations (e.g., friendship) and interdependences (e.g., two friends of a friend are also friends) among people on the formation of social relations to support better business decisions and marketing strategies.

The insights from this study can be summarized in three aspects. First, the estimated parameters of both models imply that customers' co-considerations are price and fuel economy driven because the effect networks based on vehicle price and fuel consumption are the most influential factors in forming a co-consideration link. Second, the interdependece effect arising from the shared vehicles being co-considered are found to be statistically significant in affecting the formation of a co-consideration link. Third, the model comparison study with goodness-of-fit analysis at both network level and link level demonstrates that the consideration of product interdependence using the ERGM approach helps improve the model fit.

At the end, we suggest several promising areas for future research. First, the model can be further tested to forecast products' relations and market structures using vehicle attributes data of future years instead of the training data. Also, a weighted network modelling framework can be developed to not only predict the existence of a link but also to predict the strength of the co-consideration. The weighted network models would help discover to what extent customers' consideration decisions have changed, thereby providing more concrete information to guide product design and forecast market responses.

REFERENCES

Ben-Akiva, M. E. & Lerman, S. R. (1985), *Discrete choice analysis : theory and application to travel demand,* Cambridge, Mass., MIT Press.

Carson, R. T. & Louviere, J. J. (2014), Statistical properties of consideration sets. *Journal of Choice Modelling,* 13**,** 37-48.
Chen, W., Hoyle, C. & Wassenaar, H. J. (2012), *Decision-based Design : Integrating Consumer Preferences into Engineering Design*, Springer.
Eliaz, K. & Spiegler, R. (2011), Consideration sets and competitive marketing. *The Review of Economic Studies,* 78**,** 235-262.
Frank, O. & Strauss, D. (1986), Markov graphs. *Journal of the american Statistical association,* 81**,** 832-842.
Hauser, J., Ding, M. & Gaskin, S. P. Non-compensatory (and compensatory) models of consideration-set decisions. 2009 Sawtooth Software Conference Proceedings, Sequin WA, 2009.
Hauser, J. R. & Wernerfelt, B. (1990), An evaluation cost model of consideration sets. *Journal of consumer research,* 16**,** 393-408.
Howard, J. A. & Sheth, J. N. (1969), *The Theory of Buyer Behavior,* New York - N.Y. a.o., Wiley.
Mcpherson, M., Smith-Lovin, L. & Cook, J. M. (2001), Birds of a feather: Homophily in social networks. *Annual review of sociology***,** 415-444.
Michalek, J. J., Feinberg, F. M. & Papalambros, P. Y. (2005), Linking marketing and engineering product design decisions via analytical target cascading. *Journal of Product Innovation Management,* 22**,** 42-62.
Powers, D. M. (2011), Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.
Robins, G., Pattison, P., Kalish, Y. & Lusher, D. (2007a), An introduction to exponential random graph (p*) models for social networks. *Social networks,* 29**,** 173-191.
Robins, G., Snijders, T., Wang, P., Handcock, M. & Pattison, P. (2007b), Recent developments in exponential random graph (p*) models for social networks. *Social networks,* 29**,** 192-215.
Saito, T. & Rehmsmeier, M. (2015), The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one,* 10**,** e0118432.
Sha, Z., Moolchandani, K., Panchal, J. H. & Delaurentis, D. A. (2016), Modeling Airlines' Decisions on City-Pair Route Selection Using Discrete Choice Models. *Journal of Air Transportation,* 24**,** 63-73.
Sha, Z. & Panchal, J. H. Estimating linking preferences and behaviors of autonomous systems in the Internet using a discrete choice model. 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 5-8 Oct. 2014 2014a. 1591-1597
Sha, Z. & Panchal, J. H. (2014b), Estimating Local Decision-Making Behavior in Complex Evolutionary Systems. *Journal of Mechanical Design,* 136**,** 061003-061003
Shocker, A. D., Ben-Akiva, M., Boccara, B. & Nedungadi, P. (1991), Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing letters,* 2**,** 181-197.
Shore, J. & Lubin, B. (2015), Spectral goodness of fit for network models. *Social Networks,* 43**,** 16-27.
Shumate, M. & Palazzolo, E. T. (2010), Exponential random graph (p*) models as a method for social network analysis in communication research. *Communication Methods and Measures,* 4**,** 341-371.
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**,** 071410-071410
Wang, M., Chen, W., Huang, Y., Contractor, N. S. & Fu, Y. A Multidimensional Network Approach for Modeling Customer-Product Relations in Engineering Design. ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 2015. American Society of Mechanical Engineers.
Wang, M., Huang, Y., Contractor, N., Fu, Y. & Chen, W. A Network Approach for Understanding and Analyzing Product Co-Consideration Relations in Engineering Design. DS 84: Proceedings of the DESIGN 2016 14th International Design Conference, 2016a.
Wang, M., Sha, Z., Huang, Y., Contractor, N., FU, Y. & Chen, W. Forecasting Technological Impacts on Customers' Co-consideration Behaviors: A Data-driven Network Analysis Approach. 42nd ASME Design Automation Conference, August 21-24, 2016 2016b Charlotte, NC. ASME.
Wang, P., Robins, G., Pattison, P. & Lazega, E. (2013), Exponential random graph models for multilevel networks. *Social Networks,* 35**,** 96-115.
Wasserman, S. & Pattison, P. (1996), Logit models and logistic regressions for social networks: I. An introduction to Markov graphs andp. *Psychometrika,* 61**,** 401-425.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support from NSF-CMMI-1436658 and Ford-Northwestern Alliance Project. We acknowledge Sophia Fu for her review and comments.