

## Research Article

# A Weighted Statistical Network Modeling Approach to Product Competition Analysis

Yaxin Cui <sup>1</sup>, Faez Ahmed <sup>2</sup>, Zhenghui Sha <sup>3</sup>, Lijun Wang,<sup>4</sup> Yan Fu,<sup>4</sup>  
Noshir Contractor <sup>5</sup> and Wei Chen <sup>1</sup>

<sup>1</sup>Department of Mechanical Engineering, Northwestern University, Evanston, IL, USA

<sup>2</sup>Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>3</sup>Department of Mechanical Engineering, University of Texas at Austin, Austin, TX, USA

<sup>4</sup>Insight and Analytics, Ford Motor Company, Dearborn, MI, USA

<sup>5</sup>Department of Industrial Engineering & Management Sciences, Northwestern University, Evanston, IL, USA

Correspondence should be addressed to Wei Chen; [weichen@northwestern.edu](mailto:weichen@northwestern.edu)

Received 23 June 2021; Revised 31 October 2021; Accepted 1 December 2021; Published 4 January 2022

Academic Editor: Samir Suweis

Copyright © 2022 Yaxin Cui et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Statistical network models have been used to study the competition among different products and how product attributes influence customer decisions. However, in existing research using network-based approaches, product competition has been viewed as binary (i.e., whether a relationship exists or not), while in reality, the competition strength may vary among products. In this paper, we model the strength of the product competition by employing a statistical network model, with an emphasis on how product attributes affect which products are considered together and which products are ultimately purchased by customers. We first demonstrate how customers' considerations and choices can be aggregated as weighted networks. Then, we propose a weighted network modeling approach by extending the valued exponential random graph model to investigate the effects of product features and network structures on product competition relations. The approach that consists of model construction, interpretation, and validation is presented in a step-by-step procedure. Our findings suggest that the weighted network model outperforms commonly used binary network baselines in predicting product competition as well as market share. Also, traditionally when using binary network models to study product competitions and depending on the cutoff values chosen to binarize a network, the resulting estimated customer preferences can be inconsistent. Such inconsistency in interpreting customer preferences is a downside of binary network models but can be well addressed by the proposed weighted network model. Lastly, this paper is the first attempt to study customers' purchase preferences (i.e., aggregated choice decisions) and car competition (i.e., customers' co-consideration decisions) together using weighted directed networks.

## 1. Introduction

Network modeling has emerged as a key method for statistical analysis of complex systems in a wide variety of social and engineering domains [1–5]. For example, network-based models have been applied in design collaborations [6–8], design crowdsourcing [9], new technology adoption [10, 11], and design and manufacturing systems [12]. As a complex socio-technical system, customer-product relations can be modeled with network analysis based on social network theory and techniques [13–15], where nodes

represent individual entities and links represent their relationships. Among the existing network-based modeling techniques, the exponential random graph model (ERGM) is a unique method with the ability to model the influence of both exogenous effects (e.g., nodal attributes) and endogenous effects (network configuration/nodal relations) on the network formation. In recent studies, this model has been adopted in studying customers' consideration behaviors [13], forecasting the impact of technological changes on market competitions [16], modeling customers' consideration-then-choice behaviors [17], and predicting products'

co-consideration relations [13, 18]. Particularly, the studies on how products are co-considered relate to the focus of this paper: analysis and prediction of products' competition relations.

Depending on the level of complexity, different network structures have been explored in customer-product relations, including "unidimensional" network, "bipartite" network, and "multidimensional" network, as shown in Figure 1. Among these structures, both bipartite [17, 19] and multidimensional [15] networks model customers and products as separate nodes and the relation between customer and product (customers' considerations and choices) as links. Unlike bipartite and multidimensional networks, unidimensional networks focus on product competition based on aggregated customer preference. In a unidimensional network, nodes represent products in the market, and links among them are formed based on whether customers have co-considered the products together. Prior work has shown the importance of modeling unidimensional networks for customer preference modeling. For example, Sha et al. [13] studied a binary unidimensional network to understand the influence of endogenous effects, such as the existing competition relations between car models, on the formation of new competitions in the market. Ahmed et al. [20] proposed a graph neural network approach to predict the binary unidimensional relationships between products. In this study, we adopt unidimensional network analysis to investigate product competition for two reasons. First, a unidimensional network represents competition as aggregated customer preferences and demand at the market level, from which the insights obtained would provide better decision support for enterprises than studying individual choice behaviors on competing products. Second, in a unidimensional network, customers' considerations and choices can be modeled jointly at the market level by the introduction of directed links. Therefore, it enables the prediction of market shares of different products beyond merely studying product competitions, thereby serving for the design for market systems.

Despite earlier attempts at using network models and theories in understanding the driving factors in customers' consideration and choice behaviors, existing studies have several limitations. First, the networks are simplified as binary networks, meaning that the weights or the strength of links are neglected. However, the link strength is an important aspect of understanding product competition as well as customer preferences. This is because to probe into the question of how much a competition relation between two products could be changed because of the change of designs or customer preferences, the link strength must be explicitly modeled. Second, most, if not all, past research on network models on car competition analysis does not use directed networks for modeling the final choice decision of a product but instead focus on the first stage of choice-making, that is, customers' consideration decisions. Aiming to address these limitations in the past work (as illustrated in Figure 1) [14, 16–19, 21], this study is the first, in our knowledge, to use

weighted networks as well as both choice (directed networks) and consideration (undirected networks) to study product competition and customer preferences. Figure 1 summarizes the existing studies on using unidimensional, bipartite, and multidimensional network analysis in customer preference modeling and how this work differs from them.

The new approach proposed in this study is based on the valued ERGM models that allow a link between nodes to carry weights, and such a link can be either directed or undirected. Despite the applications of network modeling techniques in different research areas, the valued ERGM technique [22] has received little attention in engineering research. Our research aims at acclimating and transferring this statistical modeling knowledge into the engineering design field for further understanding product competition relations. In a unidimensional car competition network, we study both customers' consideration and choice behaviors by establishing two types of networks as illustrated in Figure 2, an undirected network, in which links represent the co-consideration relationship and a directed network, in which a directed link between the two products co-considered indicates the customers' aggregated preferences towards the final choice decisions.

As a summary, the objectives of this research are: (a) to develop an approach based on valued ERGM to model product competition, as exemplified by the study on both weighted undirected co-consideration network and weighted directed choice network and (b) to evaluate the performance of valued ERGM in link prediction (i.e., the competition strength prediction) when nodal attributes change in different years, for example, the change of product design features when a car model upgrades from one year to another.

The primary contributions of this paper are: first, a new network-based approach using valued ERGM to explore product competition is proposed for the first time. Second, we demonstrate that valued ERGM models predict customer consideration behavior substantially better than binary ERGM models. Third, we show that valued ERGM effectively models both directed and undirected networks in analyzing aggregated customer considerations and purchasing behaviors.

## 2. Technical Background

*2.1. Exponential Random Graph Models.* Exponential random graph model (ERGM), a statistical analysis technique that serves as a formal representation of the network formation process [23], has been a popular choice in social network research. ERGM outputs a probability for every possible network that can be formed from a fixed number of nodes. This leads to a probability distribution on the set of all possible networks with the same number of nodes [23]. Mathematically, ERGMs can be expressed as a function of a set of input parameters (which can be node properties, link properties, network configuration attributes, etc.) [24], as shown in the following:

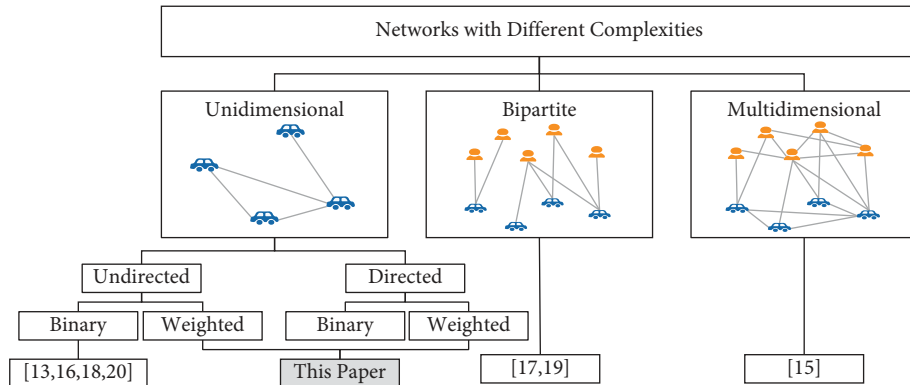


FIGURE 1: Unidimensional, bipartite, and multidimensional networks in customer-product relation modeling. We show that past work in literature did not address weighted models for directed or undirected networks. This paper fills this gap for unidimensional networks.

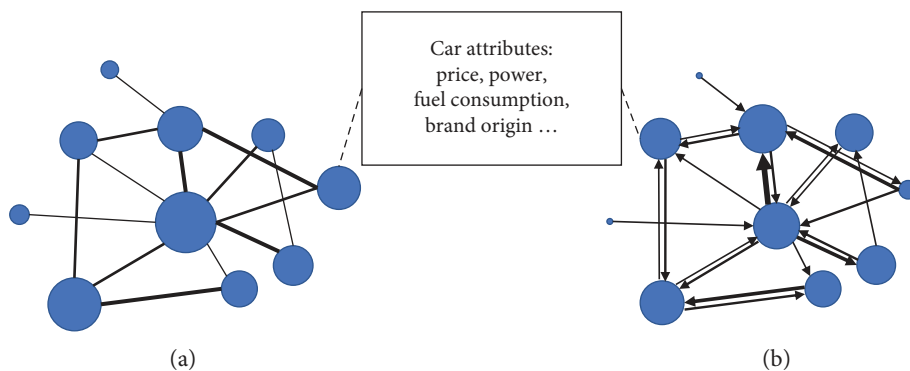


FIGURE 2: We use valued ERGM network models to study product competition in both the consideration stage (the network in (a)) and the choice stage (the network in (b)). The nodes represent cars as an example in these network illustrations and links represent competition strength.

$$\Pr(Y = y) = \frac{\exp}{(\boldsymbol{\theta}^T \cdot g(\mathbf{y}))} \kappa(\boldsymbol{\theta}), \quad \mathbf{y} \in \mathbb{Y}, \quad (1)$$

where network structure  $Y$  is treated as a random variable and an observed network  $Y$  is the network data the researcher has collected and regarded as one realization from a set of possible networks. The probability of the observed network structure is determined by network statistics  $g(\mathbf{y})$ , which can include attributes of nodes, attributes of links, and network structural attributes, along with the corresponding model parameters  $\boldsymbol{\theta}$ . As  $\boldsymbol{\theta}$  and  $g(\mathbf{y})$  are vectors,  $T$ , a transpose operator, is needed to ensure a proper dot product operation.  $\kappa(\boldsymbol{\theta}) = \sum_{\mathbf{y}' \in \mathbb{Y}} \exp(\boldsymbol{\theta}^T \cdot g(\mathbf{y}'))$  is a normalizing constant, which is a summation of the numerator over all possible networks  $\mathbb{Y}$  to make sure the function yields a realistic probability value. Equation (1) suggests that the probability of observing a specific network structure is proportional to the exponent of a weighted combination of network statistics [25]. To estimate the parameters (or learn the model from existing data), a Markov chain Monte Carlo (MCMC) procedure using maximum likelihood estimation is typically employed, and the details of the algorithm are documented in [26, 27]. The estimated parameters indicate the importance of different statistics (such as node attributes and network structural

effects) in the formation of links in a network. By analyzing the magnitude and statistical relevance of those parameters, one can find and interpret the factors that are important to the formation of the observed network.

ERGM has several advantages over traditional statistical models. For instance, unlike traditional logit models [28], they allow the interdependence among network links, which is more realistic in many network formation processes. For example, in a friendship network, when two nodes have common partners, there could be a higher possibility to have a link connecting them. ERGM also provides a flexible statistical inference framework that can model the influence of both exogenous effects (e.g., nodal attributes) and endogenous effects (e.g., the triangular network configurations that represent the three-way product competition) on the probability of forming a connection between nodes.

**2.2. Valued ERGM Model.** A limitation of traditional binary ERGM is that it cannot model networks with weighted links (e.g., the demand between two airports in an air transportation network). If one wishes to model a weighted network with the traditional ERGM, they have to first binarize the network with a link weight threshold. This dichotomization step may lead to biases and information

loss, which can eventually affect network prediction. Valued ERGM [22], a technique recently developed by statisticians, addresses this limitation by modeling the strength of links rather than merely their presence or absence. For a given set of discrete variables, a valued ERGM is expressed as follows:

$$\Pr(Y = y) = \frac{h(y)\exp(\theta^T \cdot g(y))}{\kappa(\theta)}, \quad y \in \mathbb{Y}, \quad (2)$$

where most of the parameters are the same as those in equation (1), and  $\kappa(\theta, y) = \sum_{y' \in \mathbb{Y}} h(y') \exp(\theta^T \cdot g(y'))$  also works as a normalizing constant, to make the function output a feasible probability value. Two major distinctions between the valued ERGM and the regular ERGM are the support  $\mathbb{Y}$  term and the reference distribution  $h(y)$  term.

Different from binary ERGMs, the support of a valued ERGM is over a set of weighted networks, which is often infinite or uncountable [29]. One cannot enumerate all possible weighted networks with real-valued link strengths. Thus, in a weighted network case, we need to consider what the strengths of connections are and how they are distributed. This brings in the need of specifying a reference distribution, which determines the sample space and baseline distribution of link values. The sample space is a set of possible networks given the size and density of the observed network. A reference distribution simply answers the question of what the link distribution might look like in the absence of any ERGM terms.

The ability to model valued links has greatly advanced network research as it enables researchers to conduct more nuanced examinations of network structures. Moreover, similar to traditional ERGMs, valued ERGMs are capable of modeling networks with both undirected links and directed links. Despite these benefits, valued ERGMs are still very much an exploratory area within statistical network analysis [30] due to computational difficulties.

Valued ERGMs have been employed in various applications ranging from policy studies [30] and organizational communication [31] to disease transmissions [32] and global migration [33]. An important step of using valued ERGM is to first define meaningful links and a way to measure the link strength. The definition of link strength often depends on the domain, and in the past, researchers have determined it based on factors ranging from the level of interaction between two nodes [30], the strength of friendship [31], or the total duration of human contact [32]. These links, although valued, are typically discrete in a small range such as  $\{0, 1, 2, 3\}$ . Existing methods in the social science area cannot be directly used in our study to model the valued product competition networks because: (a) the link strength in a product competition network could have a substantially large range. This infinite sample space increases the complexity of the task of prediction and (b) existing studies mainly concentrate on interpreting the models, whereas we focus on both interpretation and prediction. The prediction of the network involves network simulation based on the estimated parameters, and it can also serve as a validation of the fitted model. Despite their complexity, there are two motivations behind using valued ERGM models in this

work: (1) they can model the magnitude of competition strength between products, thereby supporting car manufacturers' strategic decisions on product positioning. As the valued ERGM will establish the functional relations between the car design features and the competition strength, the resulting model will be able to predict future market competition based on the change of certain car features, such as a design upgrade or design modification. (2) With more information captured, the valued ERGM model should demonstrate a better link prediction accuracy compared to traditional binary ERGMs.

### 3. Methodology

In a product market, the number of customers considering a pair of products ( $u$  and  $v$ ) or choosing one product over the other reflects the in-between competitive strength. To capture the product competition strength based on customers' considerations and choices, we build weighted product competition networks and model them with valued ERGMs. In this section, we outline the three main steps required for the statistical modeling of a weighted competition network:

- (1) Construct the weighted network
- (2) Train a valued ERGM model and interpret the estimated model parameters
- (3) Predict the future competition among products in the market under investigation

Our contribution is to extend the valued ERGM to modeling product competition networks. We describe the step-by-step process of building a weighted network and how to analyze it in this section. We use car design as an example in this study, but our approach can be generalized to many other product design contexts.

#### 3.1. Weighted Product Competition Network Construction.

To capture the multistage nature of a customer's decision-making process, we build two different unidimensional networks, "co-consideration network" and "choice network." The first is an undirected network that represents customers' choice set in the consideration stage, and the second is a directed network, which represents the customers' aggregated choice preferences.

In both networks, a product (in this case, a car) corresponds to a node. Each node is associated with a set of attributes such as price, fuel consumption, and engine power. We denote both networks as  $G = (V, \varepsilon, W)$ , where  $V$ ,  $\varepsilon$ , and  $W$  represent nodes, links, and weights, respectively. Figure 2 provides a simplified illustration for both the unidimensional consideration and the choice networks that we investigate. The thickness of the link between two nodes is proportional to its strength (i.e., the number of customers who co-consider the two products or choose one product over the other), and the size of the node is proportional to the popularity of the product (i.e., the number of customers who consider or purchase the product).

**3.1.1. Defining Link Strengths in the Co-Consideration Network.** In the co-consideration network, we define an undirected link between node  $u$  and node  $v$ , if there exists at least one customer who considers both cars  $u$  and  $v$  together. The number of customers who consider the two cars together is denoted as the link weight ( $w_{u,v}$ ) between nodes  $u$  and  $v$ .

**3.1.2. Defining Link Strengths in the Choice Network.** In the choice network, a directed link from node  $u$  to node  $v$  is established if there exist customers who considered car  $u$  and  $v$  together but finally bought  $v$  instead of  $u$ . The total number of customers who bought car  $v$  despite considering car  $u$  denotes the link strength from  $u$  to  $v$  and vice versa.

**3.1.3. Descriptive Network Analysis.** Descriptive network analysis helps researchers quickly explore some major characteristics of a network, such as which products are popular and how dense the network is, without going into the sophisticated statistical modeling process. It requires the computation of topological measures to assess the network structural characteristics and the implication of structural advantages [15].

The descriptive metrics adopted for analyzing a unidimensional weighted car competition network are “network weight distribution,” “centrality,” and “clustering coefficient.” The values of weights  $w_{u,v}$ , which measures the competition (while the methods we present in this paper generalize to many definitions of competition between two items, we have primarily used the term “competition” between two cars as a measure of the number of occurrences that two cars being co-considered.) strength between pairs of cars ( $u$  and  $v$ ), can be considered as a fundamental element in the weighted network analysis. The “probability distribution of weights”  $P(w)$  indicates the overall competition strength, that is, the frequency of a pair of cars being co-considered in an undirected network. We can also calculate the “centrality” of a node. In an undirected consideration network, the centrality is measured by the strength of a node, which is defined as  $s(i) = \sum_{j \in V(i)} w_{ij}$  ( $V(i)$  is the node-set of node  $i$ 's neighbourhood). It is a measure of how popular the car is. Note that as this is a measure of popularity in the consideration network, it is possible that a model is popular (i.e., considered by many people) in the consideration stage but still has a low market share. In the directed choice network, the in-strength of a node  $s_{in}(i)$  equals the sum of weights of all directed inward links, which is a measure of the popularity in the final purchase decisions. Furthermore, the fraction of a node's in-strength to the total in-strength of all the nodes ( $s_{in}(i) / \sum_{j=1}^n s_{in}(j)$ ) for directed networks represents the market share of that car model. Finally, we are interested in observing if there exist cliques of cars with intense competition. To quantify this, we use the “weighted global clustering coefficient,” which measures the overall network interconnected triplets [34]. A cluster in a weighted network is defined as a group of nodes with high-weight links between each other and with low-weight links to other nodes in the network. Therefore, a high clustering coefficient

indicates interconnected communities (car competitions within market segments) are more common in the network. While descriptive analysis provides broad insights about the network structure, it does not throw light on how different attributes quantitatively affect link formation. In what follows, we discuss the valued ERGM technique that complements the descriptive network analysis.

**3.2. Network Modeling and Interpretation.** As described in equation (2), the inputs of valued ERGMs are a reference distribution  $h(y)$  and a vector of selected input terms  $g(y)$ , such as car price and fuel efficiency, and the weighted network configurations, such as the network density.

**3.2.1. Defining a Reference Distribution.** The reference distribution  $h(y)$  acts as our prior belief about the network based on the known distribution of link weights.

Therefore, the choice of reference distribution should reflect the prior knowledge about the link strength distribution. While binomial distribution is typically used for binary networks, other choices such as the Poisson, Geometric, Bernoulli, Uniform, and Standard Normal distribution are possible for a weighted network. The exact choice of prior belief depends on the application domain and the actual data.

**3.2.2. Defining Input Variables.** Many of the variables used as input in a valued ERGM model are similar to those used in binary ERGM models, and they can be classified into three categories: network configurations, main effects, and homophily effects. Network configurations are metrics that can be used to measure network structures, such as the number of edges, triangles, stars, and degrees. The main effects correspond to nodal effects of product attributes (such as cars' price and fuel consumption), and the homophily effects are the similarity or difference between the attributes of two nodes (such as the difference of two cars' price and fuel consumption). The complete list of input variables is introduced in the case study section. Unlike any dyad-independent binary ERGM statistics expressed as  $g_k = \sum_{(i,j) \in Y} x_{k,i,j} y_{i,j}$ , where  $x_{k,i,j}$  denotes to (nodal/edge) covariate and  $y_{ij}$  is allowed to have values either 0 or 1, in the valued ERGM,  $y_{ij}$  has a larger range of choices (0,1,2,3, . . . in our case). As for the network configuration terms, valued ERGM can handle network sparsity, mutuality, individual heterogeneity, and triadic closure via various network structural terms [22].

**3.2.3. Interpreting Valued ERGM Parameters.** The result of the valued ERGM is a set of estimated coefficients and associated  $P$ -values for all variables. Network configuration effects indicate link independence, that is, the formation of links due to the presence of other links [35]. The estimation of those effects can be seen as evidence of the prevalence or absence of certain structures (such as density, transitivity, and star effects) in a network. For example, a negative estimate of “edges” indicates that the competition network has

a low density. The impact of main attributes refers to how an attribute might influence a product's propensity to form a link in a co-consideration network or a choice network. For the car example, we examine the selected car attributes, and the result will help designers understand whether cars with certain attributes, for example, a higher price and lower fuel consumption, are more likely to be considered by customers and win a competition. The homophily effects test the hypothesis that cars with more similarities in different attributes are likely to be co-considered, which is a common explanation established in social relations.

**3.3. Market Competition Prediction.** While statistical network models are typically used to interpret what factors lead to link formation or dissolution, predicting what a network will look like in the future is useful for manufacturers to make strategic decisions. In practice, if manufacturers can predict how the competition between car models would change when certain product design attributes are changed, they can use this knowledge to position their products in the market strategically against competitors. Using the estimated parameters of input variables in valued ERGM, we can predict competition networks in the future, with new car attributes as input.

Based on the valued ERGM equation (equation (2)), the distribution of network models is determined by a base network structure, estimated parameters, input variables, and a reference distribution. Therefore, when predicting a future competition network, we substitute the old car attributes with new ones and derive the distribution of the predicted network structures based on the valued ERGM formula. Then, we draw many samples from the network distribution (simulated networks) and take the averaged network structure as the aggregated network, which represents the central tendency (highest probable network) of all simulated networks. We use this aggregated network as our prediction and compare it with the known network in the future to show our model's accuracy.

Future predictions using aggregated simulations can be made for either the co-consideration network or the choice network. In the predicted co-consideration networks, the number of competitors and their strengths are predicted. In the predicted choice networks, the manufacturers will get an understanding of which car models are their main competitors. In the next section, we show how the methods and the process discussed above are applied to two real-world vehicle data sets.

## 4. Case Studies

In this section, we demonstrate the use of the valued ERGM approach to study the Chinese car market. We use data from a new car buyer survey as a test example. Weighted network modeling can be applied to different stages of decision-making of a customer, which corresponds to different types of network models (undirected and directed). We show two case studies covering different aspects of network structures and the decision-making process of customers. The first case

study focuses on the initial stage of customer decision-making and uses an undirected co-consideration network model. The second case study focuses on the final stage of choice-making using a directed choice network model. In this case, one of the cars that are co-considered by the customers "wins" the competition, thus is finally purchased.

**4.1. Data Description.** Our data set contains survey data from 2013 to 2014 in the China market. In the survey, there were around 53,000 and 60,000 respondents, respectively, in 2013 and 2014, who specified which cars they purchased and which cars they considered, before making their final choice. Each customer indicated at least one and up to three cars which they considered. The data set also contains many car attributes (e.g., price, power, brand origin, and fuel consumption) and customer-specific attributes (e.g., gender, age, etc.).

**4.2. Case Study 1: Car Co-Consideration Network.** In this case study, we use valued ERGM models to study the competition between any pair of car models reflected by the number of co-considerations received between them.

**4.2.1. Step 1: Network Construction and Characterization.** To study car co-consideration, we start by creating a car co-consideration network based on customers' survey responses in the 2013 survey data. For purpose of validation, we control the studied market size and a random sampling of 50,000 customers was made. It is noteworthy that customers who have only considered one car in the survey are removed because they do not provide valuable information about product competition, and our network currently has taken roughly 38,000 customers. The network consists of 296 unique car models as network nodes. The link between a pair of nodes carries the weight equal to the number of customers who considered both the car models together in their consideration set. The overview of the 2013 co-consideration network is shown in Figure 3. As the node size is proportional to the weighted degree of a car model, a larger node size depicts a more popular car model because it is considered by more customers. Similarly, a thicker link width displays a stronger co-consideration relationship (competition) between a pair of cars. Figure 3 also shows a glimpse of a three-way competition. In this example, cars "Great Wall Hover" and "Honda Dongfeng CRV" appear together in the consideration set of 18 customers in 2013 and 30 customers in 2014, showing that their competition has potentially increased in one year (note the sampled market size for 2013 and 2014 are the same). In contrast, cars "VW SVW Tiguan" and "Honda Dongfeng CRV" appear together in the consideration set of 201 customers in 2013 and 192 customers in 2014. This shows that their competition has decreased in one year, although both car models are still more popular than the "Great Wall Hover," as indicated by the sum of all link strengths connected to them.

Table 1 presents a summary of our network's descriptive characteristics. Network density, which calculates the

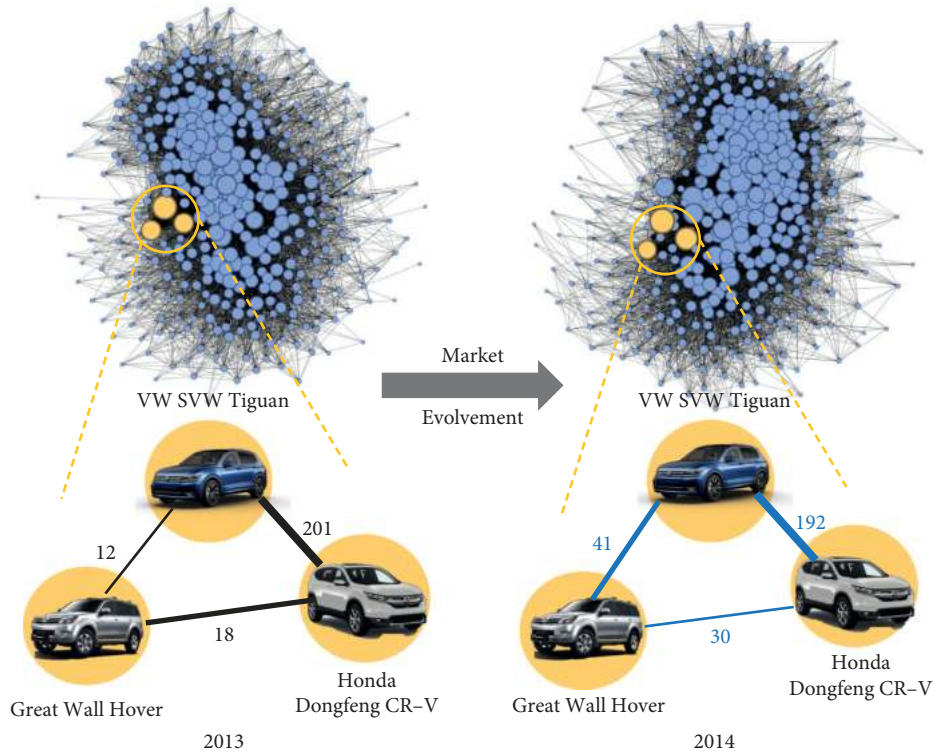


FIGURE 3: An overview of the 2013 and 2014 co-consideration network. (a) The blue nodes represent car models and black links represent co-consideration relations. The node size is proportional to the weighted degree of a car model, and the link width is proportional to the link strength of the co-consideration relation. And, an example of the local co-consideration network between three cars changing from 2013 to 2014 (b).

TABLE 1: Summary of 2013 co-consideration network descriptive characteristics.

No. of nodes	Network density	Ave. strength	Ave. degree	Ave. weighted degree	Global clustering coefficient
296	0.152	5.323	22.355	118.80	0.616

portion of the potential connection between all nodes that are actually connected in a network, shows that among all possibly connected car models, 15.2% of them are being co-considered, and an average of 5.323 customers consider any connected car models indicated by the average strength. The average degree means that each car competes with 22.355 cars on average. The average weighted degree indicates a car is co-considered with other cars by 118.80 customers on average. The average global clustering coefficient of 0.616 suggests that car models are very likely to engage in a multiway competition.

To build the valued ERGM network model, we select the set of most representative car attributes based on the selection criterion used in a previous study [13], including price, engine power, fuel consumption, market segment, import status, and car make origin. This selection allows us to use our prior work as a baseline for comparison purposes. We apply log transformation to price (in Chinese Yuan RMB) and engine power (in brake horsepower BHP) to normalize the range of attribute values and reduce the large outlier effects. Fuel consumption is calculated by the ratio of consumed gasoline (in liters) to driving distance (in 100 km), and a smaller fuel consumption value speaks for higher fuel

efficiency. The market segment is a categorical variable that contains 17 car segment codes provided by Ford. The variables of import and make origin are related to the car's brand information, and 35.1% of cars are imported from Europe, the United States, Japan, and South Korea, and 64.9% of cars are domestically produced in China.

*4.2.2. Step 2: Network Modeling and Interpretation.* In the implementation of the valued ERGM model, we assign the selected car attributes to network nodes and the occurrence of co-considerations to the link strengths. Based on the sample space of link strength (nonnegative, integer, and not bounded), the available reference distributions are Poisson distribution and geometric distribution. In an empirical setting, Poisson distribution provides a converged and legitimate result; therefore, we have chosen Poisson distribution as the reference distribution.

The input variables can be divided into three categories: the network configuration effects, the main effects [13], and the homophily effects. The whole set of input variables can be found in Table 2. We use the statistical network analysis package "Statnet" in R programming, in which the valued

TABLE 2: Estimated coefficients of the 2013 co-consideration network for a weighted network and binary networks.

Input variables	Weighted	Binary 1	Binary 2	Binary 3
<i>Network configuration effect</i>				
Sum/intercept	-9.54***	-7.24***	-8.80***	-11.72***
<i>Main effect (nodal attributes)</i>				
Import	-1.46***	-1.00***	-1.31***	-1.59***
Price (log2)	0.27***	0.18***	0.21***	0.25***
Power (log2)	0.05	0.07	0.01	0.04
Fuel consumption	-0.03***	-0.06***	-0.02	-0.01
Brand origin (USA)	1.42***	1.17***	1.43***	1.72***
Brand origin (Europe)	1.11***	0.82***	1.09***	1.31***
Brand origin (Japan)	0.45***	0.46***	0.58***	0.67***
Brand origin (Korean)	0.75***	0.69***	0.90***	1.07***
<i>Homophily effect (dyadic attributes)</i>				
Market segment matching	1.16***	0.76***	0.93***	1.10***
Brand origin matching	0.87***	0.82***	0.97***	1.13***
Price difference (log2)	-1.90***	-1.35***	-1.83***	-2.25***
Power difference (log2)	-0.06	0.12	0.04	-0.12
Fuel consumption difference	-0.30***	-0.26***	-0.38***	-0.41***

Note: \*\*\*  $p < 0.001$ .

ERGM is integrated [36]. The second column of Table 2 (i.e., “Weighted”) shows the estimated coefficients from fitting the valued ERGM models. The sum/intercept variable serves as a constant term in valued ERGM, and it estimates the likelihood of two cars’ co-consideration strength without any knowledge about the cars’ attributes. All the input variables, except the main effect of power and the homophily effect of the power difference, are statistically significant at the level of significance of 0.05. As all variables are normalized to a similar order of magnitude, the differences in the coefficients denote their relative importance in the model fit. Among the main effects, the coefficient of import effect is negative, but the coefficients of brand origin from different countries are positive. This implies that customers tend to consider domestically made cars with foreign brands, such as Ford Changan Focus, Honda Dongfeng Civic, and so on. Variables such as price, power, and fuel consumption are not as important as the other main effects. We observe that the coefficients corresponding to the homophily effects are mostly positive and significant. This indicates that the homophily effects may play an important role in forming the competitive relations between two car models, which verifies our common beliefs. Among the homophily effects, market segment matching and brand origin matching are significant. This may reveal that car models within the same market segment and the same brand origin tend to be co-considered by customers. Furthermore, a statistically significant large negative coefficient of price difference shows customers prefer to consider cars in a similar price range. This observation aligns with our intuition, as a customer may consider cars within his/her budget range.

To validate the findings of our valued ERGM model, we compare it against binary ERGM models (where the network only considers the existence of the competition instead of the competition strength). We set the model terms, such as the attributes considered and MCMC termination criteria (i.e.,

the  $P$ -value of Hotelling’s  $T^2$  test for equality of MCMC-simulated network statistics exceeds 0.5 [37]), to be the same between all models. However, as the performance (as measured by network prediction) of the binary ERGM model is sensitive to the method of binarization, we compare our method with multiple binary networks. Specifically, we add results for three cutoff link strength values, that is, 1.0 (the first quantile), 2.0 (the median), and 4.0 (the third quantile), and we denote the link strength larger than the cutoff value as a link in the binary network. The model results are shown in Table 2.

The estimations of different binary network models show that the effects of fuel consumption and power differences become nonsignificant when the cutoff value changes, which indicates that the different choices of the cutoff value for the binarized network can result in different model results and different interpretation of the model. When using valued ERGM, this inconsistency is not observed due to the lack of need for binarizing the network with an artificial cutoff value.

However, while binary networks with different cutoff values have some inconsistency among themselves, it is important to understand that overall the ERGM modeling package for binary and valued networks seems reliable for our problem. This is evident due to two reasons. First, the relative value of attribute coefficients in different binary models does not drastically change. Second, we see that most coefficients have a similar sign and relative value between valued and binary networks too, which shows the ERGM models are reliable in estimating the attribute effects and converge to similar values.

The valued ERGM provides a few new insights. For instance, the coefficient of market segment matching is larger than brand origin matching for valued ERGM in contrast to binary ERGM, which implies brand origin matching plays an important role during the link formation but market segment matching is more dominant in creating higher competition strength. We trust these interpretations



more, as the valued ERGM also has better network prediction performance.

To test the consistency of the model with different customer samples, multiple random samples have been run to sample 50,000 customers. The estimated parameters are stable and consistent across varied sampling results in 50 random samples. The mean and standard deviation of the estimated parameters in different samples are shown in Table 3.

*4.2.3. Step 3: Validation and Prediction.* We perform three different types of validation to examine both the model fit and the predictive power, as elaborated in the following.

*(1) Trained Model Prediction Matches the Link Strengths in the Training Data Pretty Well.* We start the model validation by performing simulations with the current network configurations and the estimated coefficients of the selected model terms. More concretely, we create 100 simulated networks with the 2013 car co-consideration network configurations and the estimated parameters in Table 2 and then take the average of the link strength values from 100 simulations and denote it as the aggregated simulated car co-consideration strength. The comparison of the link strength between the simulated network and the original network reveals the goodness of the model fit. Figure 4(a) plots the link strengths of the true network compared to the aggregated simulated network along the diagonal. We observe that two sets of link strengths are positively correlated, where a perfect  $y = x$  line indicates a perfect fit. This is manifested by the Pearson coefficient of 0.988 and the coefficient of determination ( $R^2$ ) of 0.976.

*(2) Trained Model Predicts Link Strengths of Future Unseen Data Reasonably Well.* In practice, the benefit of training a statistical model is to predict the future state and behavior of networks that are unseen. While the market competition between different car models varies yearly, we test whether our fitted co-consideration model can be utilized to predict the co-consideration relationship in the future market. Figure 3 illustrates an example of the real market evolution. It can be observed that in 2014, Great Wall Hover gains more customers' consideration, and the strong co-consideration relationship between VW Tiguan and Honda CRV decreases slightly. Our examination of the model's predictive power uses a similar method of network aggregation as used in the above validation study but with the input of 2014 car attributes as the updated node attributes. With a similar simulation process, we derive the aggregated predicted co-consideration network for the 2014 market data and compare it with the actual co-consideration network. The scatter plot of the actual link strength and the predicted link strength is reported in Figure 4(b), with a  $R^2$  of 0.794 and a Pearson coefficient of 0.893. More importantly, we observe that although there exist some deviations between the prediction and the true link strength in the lower range of the link strength values, the prediction is better when the link strength is larger. In practice, the ability to correctly

TABLE 3: With 50 different samples of 50,000 customers, the mean and standard deviations of estimated parameters show the model is stable and consistent.

	Mean	Standard deviation
Sum	-9.702 82	0.061 031
Import	-1.471 65	0.004 381
Price	0.281 495	0.002 867
Power	0.046 589	0.004 776
Fuel consumption	-0.032 57	0.001 119
Brand origin (USA)	1.410 614	0.003 135
Brand origin (Europe)	1.101 363	0.002 585
Brand origin (Japan)	0.449 603	0.002 032
Brand origin (Korean)	0.751 167	0.004 675
Market segment matching	1.152 023	0.002 034
Brand origin matching	0.874 848	0.001 701
Price difference	-1.896 83	0.002 395
Power difference	-0.072 07	0.007 694
Fuel consumption difference	-0.298 12	0.001 078

predict large link strength values is more important because they indicate more intense competition where major players in the market are always involved. Moreover, the model's performance is robust and insensitive to the changes in the estimated coefficients. For example, on changing the coefficients by  $\pm 10\%$ , the change in  $R^2$  is always less than 2%.

*(3) Valued ERGM Has Higher Precision and Recall Compared to the Baseline Binary Models.* We want to further compare the prediction results with the previous binary nonweighted network baseline. However, for comparison, we have to convert a simulated weighted network to a binary counterpart using a cutoff value of the link strength.

We choose three different cutoff values, 1.0, 2.0, and 4.0, for creating the binary network. These cutoffs are determined based on the first, second, and third quantiles from the actual network link strength distribution. After that, we compare the predicted co-consideration network with the actual binary network. This comparison allows us to measure the false and true positive rates as metrics to evaluate the model performance. More specifically, we draw the receiver operating characteristics (ROC) curve for each cutoff value. ROC curve [38] is a performance measurement for classification problems at various threshold settings of the predicted probability, and the larger the area under the curve (AUC) is, the better is the model's predictability.

As shown in Figure 5, for all the ROC curves, AUC for the weighted network is larger, which indicates a better predictive performance of valued ERGM compared to binary ERGM. As the cutoff value increases, the performance of binary ERGM keeps, while the performance of valued ERGM becomes better and better. This is because as the binary network becomes sparser, only links with higher strength are preserved and valued ERGM has better performance in predicting those links.

*4.3. Case Study 2: Crossover SUV Choice Network.* In this case study, we use valued ERGM models to study the competition between crossover SUV cars in the final choice stage of a customer.

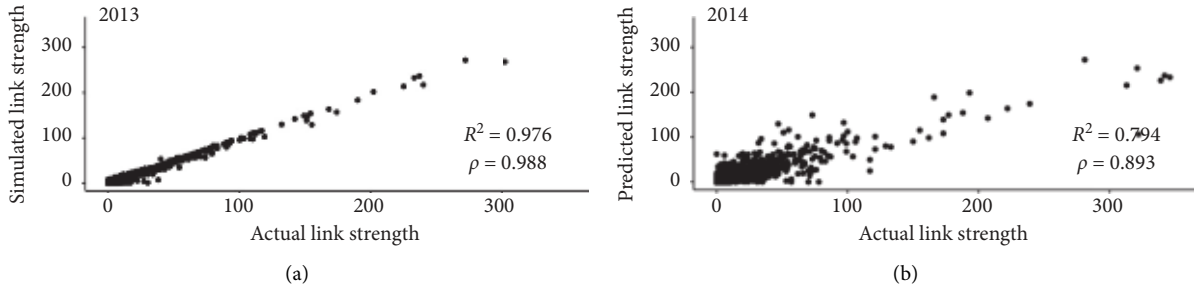


FIGURE 4: The goodness of fit using link strength comparison between the trained network and simulated network: (a) link strengths of the trained network with the aggregated simulated network for 2013 and (b) link strengths of the true network with the aggregated simulated network for 2014 (unseen future data).

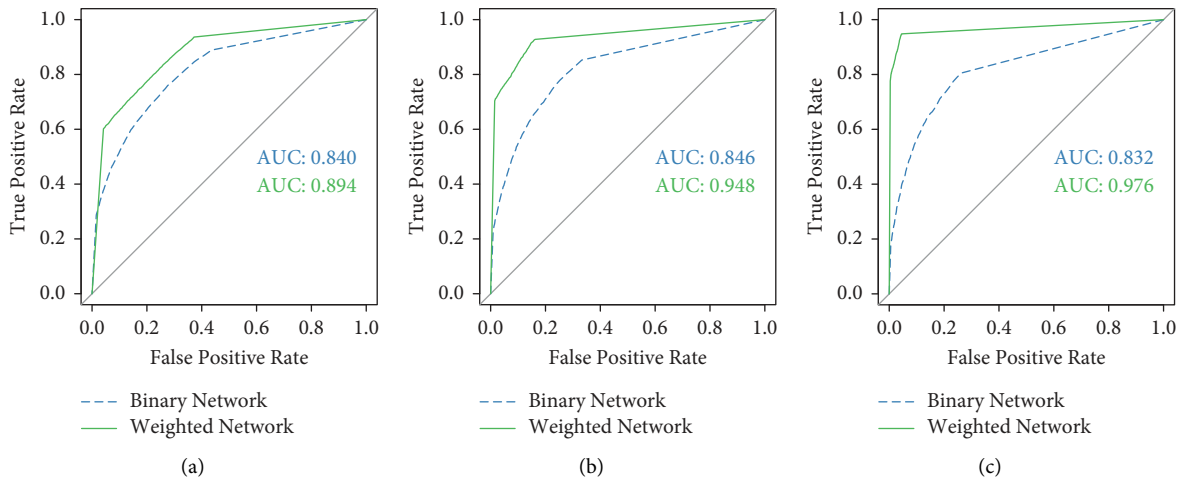


FIGURE 5: Receiver operating characteristics (ROC) curves comparing the valued ERGM model with binary ERGM models with different cutoff values for network binarization on the 2014 car competition network. We observe that irrespective of what cutoff value is used, valued ERGM models have higher precision and recall values than other models: (a) cutoff = 1.0, (b) cutoff = 2.0, and (c) cutoff = 4.0.

**4.3.1. Step 1: Network Construction.** In the second case study, we focus on the market competition among crossover SUVs, such as Ford Edge and Mazda CX-7, which are designed with the body and space of an SUV but the platform of a sedan. This type of car models has gained increasing attention in recent years and has witnessed considerable growth in many countries, owing to the low cost, compact size, stylistic design, and better maneuverability. There are 14 crossover SUV models in the 2013 survey data, and we have collected all survey data of which customers have either considered or chosen a crossover SUV model in that year. This gives a total of 1,975 customer observations (different from case study 1 when 50,000 customers are selected from the entire market, we focus on the crossover SUV car segment and only select the customers who have considered or purchased a crossover SUV model.). The directed choice network is established based on the customers' purchase behavior as described in the previous section, and all competitors in the network are divided into four segmentation groups: Sedan, SUV, luxury or sport, and crossover SUV. The visualization of the choice network is plotted in Figure 6, where the node size of a crossover SUV reflects the number of customers who have purchased it.

Overall, there are 217 car models in the crossover SUV choice network. All the links are directed and point to the “winner” in a competition. The average link strength is 2.431 corresponding to the average number of customers' purchases among all co-considered cars. A unique feature of the choice network is that the in-strength of a node is correlated with its market share.

**4.3.2. Step 2: Network Modeling and Interpretation.** The procedure of network modeling of a choice network shares many similarities with that of a co-consideration network using the valued ERGM approach. However, as the choice behavior is not symmetric between pairs of nodes, the model terms are further specified for inward nodes or outward nodes. Specifically, the main effects in Table 4 refer to the nodal attributes of the inward nodes; hence, we can learn the important attributes of the “winners” and find possible reasons behind the popularity of a car model. Besides, we have added two network structural effects, “cyclical weights” and “transitive weights,” which measure the triadic closure and refer to the links from  $i \rightarrow j$  that have two paths (two paths refer to a network structure that there are two edges connected from  $i$  to  $j$ :  $i \rightarrow h \rightarrow j$ ) from  $j \rightarrow i$  and from

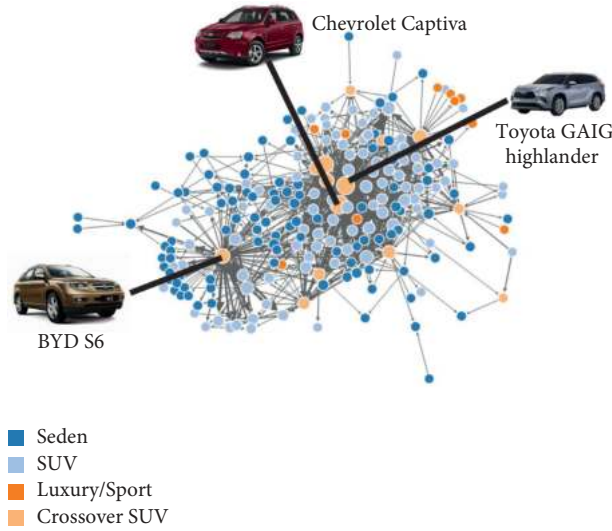


FIGURE 6: A force-directed graph visualization of the 2013 choice network for crossover SUVs. We observe that most crossover SUVs compete with Sedans and SUVs.

TABLE 4: Estimated coefficients of the 2013 choice network.

Input variables	Model 1	Model 2	Model 3
<i>Network configuration effect</i>			
Sum/intercept	6.19***	6.11***	5.77***
Cyclical weights		-0.06**	
Transitive weights			0.16***
<i>Main effect (inward node attributes)</i>			
Import	0.03	0.03	0.03
Price (log2)	-1.05***	-1.04***	-1.03***
Power (log2)	0.30*	0.30*	0.30*
Fuel consumption	0.58***	0.58***	0.57***
Brand origin (USA)	0.82***	0.81***	0.78***
Brand origin (Europe)	0.15	0.13	0.15
Brand origin (Japan)	0.80***	0.80***	0.75***
Brand origin (Korean)	0.56***	0.56***	0.54
<i>Homophily effect (dyadic attributes)</i>			
Market segment matching	0.68**	0.69***	0.67***
Brand origin matching	0.99***	1.00***	0.98
Akaike information criterion (AIC)	-68 209	-68 211	-68 252
Bayesian information criterion (BIC)	-68 113	-68 106	-68 147

Note:  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , and \*\*\*  $p < 0.001$ .

$i \rightarrow j$ , respectively (Figure 7). More precisely, in the product competition market, it accounts for a hierarchical three-way competition. The cyclical weights refer to the case when customers prefer car  $k$  to car  $j$  and prefer car  $i$  to car  $k$  while preferring car  $j$  to car  $i$ . The transitive weights refer to the case when customers prefer car  $k$  to car  $j$  and prefer car  $i$  to car  $k$ , while preferring car  $i$  to car  $j$ .

Table 4 shows the estimated coefficients from fitting three directed valued ERGM models with different model terms. The first model is a baseline model with main effects and homophily effects, and the second and the third models include network structural effects to further investigate the endogenous network effect influence. Among all three networks, the estimated coefficients are consistent with small variations. In the choice network, the car models with

lower prices, higher power, and higher fuel consumption are more likely to be bought by customers. This result is consistent with our common sense. Please note that for the group of customers who have a preference for crossover SUVs, they possibly prefer a model with higher fuel consumption, which is usually in a company with a higher power. Meanwhile, imported cars are not always preferred by this survey population, but a car with foreign brands still shows a positive effect on customers' final choice. Furthermore, the homophily effects have significant positive effects on the choice decisions, and the underlying reason is similar to the first case study. Also, in models 2 and 3, the cyclical weights have a negative effect, while the transitive weights have a positive effect. This implies that in a three-way competition, the competition relations tend to be

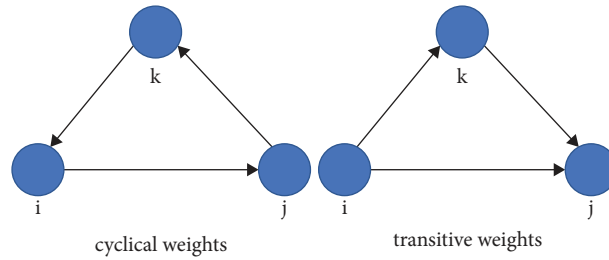


FIGURE 7: An illustration of cyclical weights and transitive weights. It refers to three-way competition in the market: (a) cyclical weights and (b) transitive weights.

transitive, meaning that if car A “wins” a competition over car B and car B “wins” a competition over car C, then car A is likely to “win” car C. Therefore, it can be inferred that the directed network market is hierarchical. We have also reported Akaike information criterion (AIC) and Bayesian information criterion (BIC) values for three models, a lower AIC and BIC value indicates a better model fit [39], and the models with network configuration statistics fit slightly better than the baseline model, which indicates that those network configurations could play an important role in the competition network formation.

**4.3.3. Step 3: Validation with Pairwise Competition Comparison.** We validate our model using two methods: (1) predicting pairwise competition and (2) estimating the market share of each car model. We first evaluate the model fit at the level of pairwise competition. Given the original network structure, one can identify the “winner” in each pairwise competition by counting the customers’ choice prevalence. For example, among 25 customers who have considered both cars A and B, 15 customers bought car A, and 10 customers bought car B, and then car A is denoted as the “winner” in this “A-B” competition. After generating the simulated choice networks based on the fitted model, the aggregated (i.e., averaged) link strength is used to quantify the pairwise competition. The results show that the simulated choice networks obtained from three different models can correctly predict over 60% of the pairwise competitions (i.e., 62.10%, 64.08%, and 61.19%, respectively).

**(1) Validation with Market Share Comparison.** In a directed choice network, the in-strength of node  $s_{in}(i)$  is related to its market share. Hence, we can further validate the choice network by comparing the simulated market share for each crossover SUV with its true market share. Specifically, the in-strength fraction  $s_{in}(i) / \sum_{j=1}^n s_{in}(j)$  is calculated based on an observed choice network for the actual market share of the crossover SUVs. Then, the simulated market share is derived by averaging the in-strength of the nodes from 100 simulations. The comparison of actual market share, simulated market shares of three different models, and the uniform market share (which assumes all crossover SUVs have the same market share and serves as a baseline) is plotted in Figure 8. Even though there exists discrepancy for some car models (e.g., Mazda CX-9 and GM USA Buick Enclave), most of the predictions of car models show a consistent

trend with the actual market share. Compared to the baseline of uniform market share, all simulated market shares have a  $R^2$  value above 0.7, which indicates that more than 70% of the observed variation can be explained by the fitted choice network model. Among them, model 1 has a  $R^2$  value equal to 0.77; model 2 has a  $R^2$  value equal to 0.70; and model 3 has a  $R^2$  value equal to 0.74. As a side note, the models adding more network attributes do not provide a better-simulated market share than the baseline model (model 1), which could be raised by the sparsity and less influence of the network structure.

While valued ERGM shows a reasonably good fit for the relative pairwise competition and the market share, it does not predict well the absolute value of weights in the choice network. This is true in predicting both the current market and the future market. We suspect that this is due to the sparsity and directionality of the network. The network constructed in this case study only contains crossover SUVs, thus leading to a very low network density of 0.02.

## 5. Discussion

While the valued ERGM model provides many advantages over existing statistical models, it is a relatively new model with a few theoretical and practical challenges that require attention and more research. In this section, we summarize the benefits and limitations of the valued ERGM models and discuss how they pave the path to future research directions.

**5.1. Supporting Engineering Design Decisions Using Valued ERGM.** One of the goals in using the valued ERGM model is to demonstrate how the approach helps identify the important factors that influence product competition. These factors can support stakeholders in making strategic decisions. However, it is important to note that while the theoretical model allows one to estimate the importance of any attribute, the analysis in specific case studies may also depend on what product data is available and whether there indeed exists any relationship between product attributes and customers’ choice decisions. To understand this, let us consider three hypothetical situations. In the first situation, a customer decides to buy a car merely based on the size of the car engine. Using a valued ERGM model, the analysis results show that the size of the engine (or power that is correlated with it) has a significant positive coefficient. In such a case, the network models inform that increasing the engine size can help gain a larger market share. However, increasing the

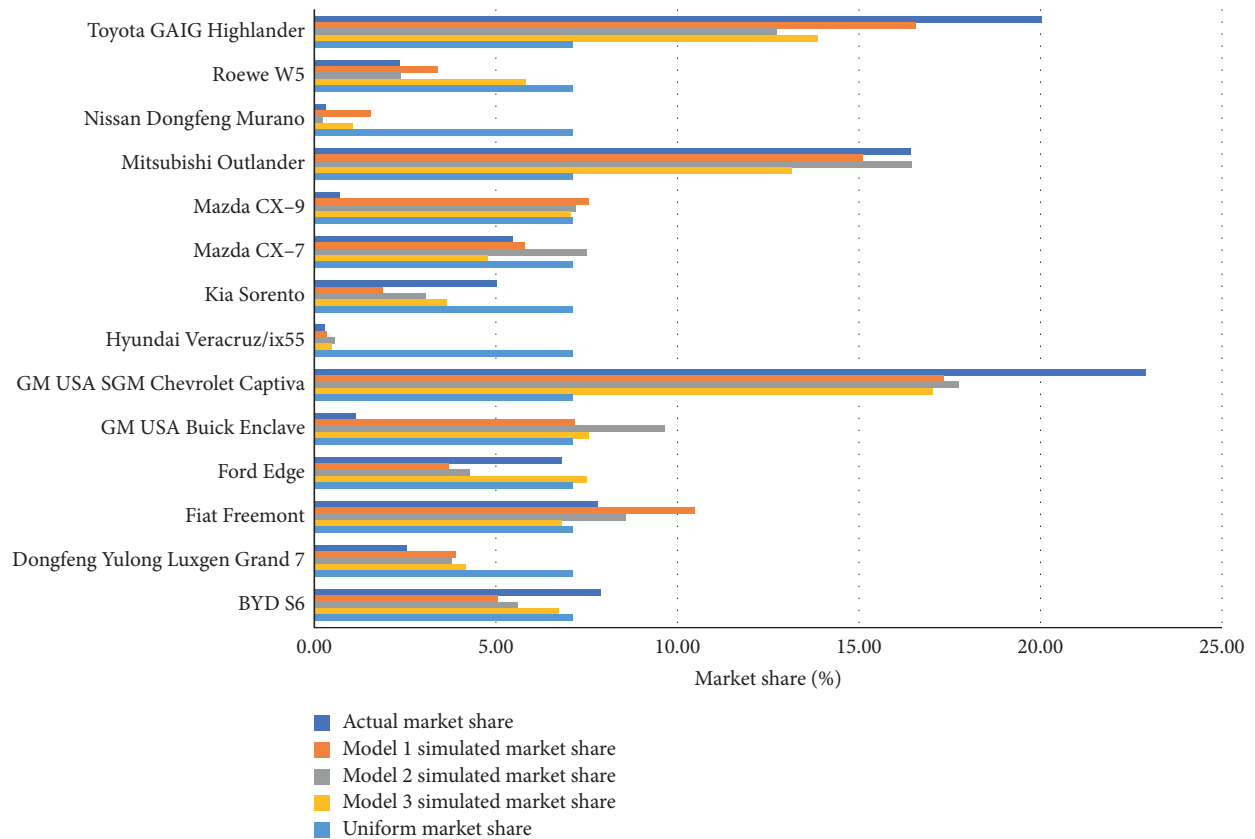


FIGURE 8: Valued ERGM prediction of 2013 crossover SUVs market share aligns with the true market share.

engine size will inevitably increase the manufacturing cost, thus leading to a higher price. This, on contrary, may negatively influence the market share. There is obviously a trade-off decision the car manufacturer has to make; then the network model should help car manufacturers make decisions of choosing the right combination of design features.

In the second situation, we assume that a customer decides to buy a car merely based on the quality of its air-conditioning (AC) system. If the data we analyze does not include the AC design attribute, the results will not be able to provide specific insights into the impact of AC design on customers' choice behaviors. The only remedy for this is to collect data that captures the relevant attributes for the choice analysis. In the third situation, we assume that customers' choice behaviors are only influenced by social and/or cultural factors but not car design features. In such cases, the coefficients of all design attributes may not have statistical significance. This indicates that improved design features may not help automakers gain more market share. Hence, the guidance provided to the manufacturer is to not waste resources on improving factors that do not have an impact.

In this paper, the customers' choice behaviors described in the two case studies are a mixture of the three situations. For example, we find that some design attributes have a statistically significant influence, but we also discover that this data set lacks information about certain car design attributes. Finally, many design attributes studied are not

statistically significant, indicating that those attributes may not play a role in customer decisions.

From our current results for both case studies, we successfully identify a few factors that impact engineering design decisions for product consideration. Specifically, in the co-consideration network of case study 1 (Table 2), we observe that a car designer may want to reduce fuel consumption (which relates to engine efficiency) to increase the competitiveness of their car models. Although factors such as price, power, and fuel consumption are statistically significant, they do not directly provide actionable design guidance for a car manufacturer. In the choice network of case study 2 (Table 4), the model results help decision-makers with strategic planning. For example, in the crossover SUV market, the improvement of fuel consumption may not increase the likelihood of a car being purchased. Instead, reducing the price and increasing the power could be helpful improve the market share. We notice that our data set lacks certain car design attributes that may be influential to customers' choices. In future work, we aim to address this issue using crowdsourced data inputs. Moreover, we have discovered the important effect of particular network configuration statistics, such as "cyclical weights" and "transitive weights" in Table 4 in the choice network. It manifests the advantages of (valued) ERGMs in utilizing network configurations to capture endogenous effects in a market. The insights into these endogenous effects help car manufacturers gain an in-depth understanding of the market and their competing opponents.

**5.2. Trade-Off between Feature Engineering and Model Interpretability.** In valued ERGM models, we start with a large collection of features. These features can be node-specific (e.g., car fuel efficiency and price), link-specific (e.g., the price difference between two car models), or network-specific (e.g., popularity and density). The choice of what features to use has a large impact on the goodness of fit of the model, the estimated coefficients as well as their statistical significance. While we use automated methods for feature selection (which largely select features that are uncorrelated), the process is often manual. In contrast, one can use modern deep learning models to learn hierarchical feature representations. Yet the deep learning models are largely black box and are hard to interpret, which is one of the key reasons for us to adopt the interpretable and theory-grounded statistical network models in this study. In the future, we will attempt to find the middle ground of reducing dependence on feature selection, while still retaining model interpretability by combining the two methods.

**5.3. Numerical Issues with Valued ERGMs.** Existing literature reports two numerical issues of (valued) ERGMs: the reliability of model interpretation and computation issues for large networks.

**5.3.1. Reliability.** In recent years, there have been critiques of using (valued) ERGM packages related to the accuracy of inference methods reported by the statistical software for ERGM. While some experiments suggested that the variants of ERGM models can work well even with a relatively small sample taken from the network [40], Shalizi and Rinaldo [41] have argued that ERGMs are designed for modeling the entire network. In many applications, the data used consists of a sampled sub-network, which could lead to inconsistency of interpretation due to the MCMC sampling process. However, our first case study is unlikely to suffer from the reported issues due to two reasons: (1) the subset of customers in our network for the first case study only changes the link strength magnitude and we still use all nodes and (2) we also test with different subsets samples of customers and find that the results are similar, which indicates the reliability of our network models. For the second case study, we use a particular market segment of cars to create the network, which may suffer from reported limitations. Hence, we are cautious in generalizing our findings from the study on crossover SUVs to other car segments.

**5.3.2. Computation Issue for Large/Complex Network.** It is reported in the literature [42] that for large and complex network structures, the MCMC approach to estimate ERGM parameters may not converge. In our work, this limitation can be a problem for some stakeholders. There is some recent work on developing scalable binary ERGMs [43, 44], and the extension of such methods to valued ERGMs can help alleviate the scalability problem for large data sets. Another approach that can improve the scalability of valued ERGMs is to use kernelized approximate Bayesian

computation. It can improve computational efficiency and is being adopted by popular packages [45] as an alternative to MCMC.

## 6. Conclusion

In this paper, we enhance the network modeling approach for analyzing customer preferences and product competition by viewing customer-product relations in the context of a complex socio-technical system. With a focus on the unidimensional network as the aggregated result of customer preferences and the social and market environment, we exhibit how valued ERGM models can be used to model directed and undirected product competition networks with nonbinary link strengths. The method enables designers to estimate the major factors that affect customers' consideration and choice behavior and that can help predict the strength of future market competition when a manufacturer changes some product attributes.

This work has three main contributions. First, we extend the newly developed valued ERGM, which has traditionally been confined to social network modeling, to study competition between products. This network modeling approach enriches the knowledge base of product design modeling techniques. Second, by developing a procedure of weighted network construction, interpretation, and validation, we demonstrate that valued ERGM models provide a better model than binary ERGM, as measured by model fit and prediction accuracy for car competition. Third, this paper is the first work to study aggregated purchase preferences using a "directed" unidimensional network. The directed network we create is unique, as it encodes information from two stages of decision-making, both the final purchase decision as well as the items considered by the customers.

The case studies in this paper show how network models are used to systematically analyze large real-world networks. For the first case study, which analyzes the co-consideration competition between 296 cars, we show that homophily effects, affecting the differences between two cars, are more important than the main effects in predicting link strength. Cars are generally found to compete more with other cars from the same market segment, same brand origin, and similar price range. In the second case study, which focuses on the crossover SUV market, we analyze a network of 217 cars and find that cars that are considered by more people are also purchased more often. In future work, we aim to analyze how valued ERGM can help study new domains and further investigate ways to integrate feature learning methods such as deep learning with valued ERGM models while retaining their interpretability. Improving the scalability of these models to larger data sets and using them for dynamically evolving car competition is another interesting avenue of research.

## Data Availability

Anonymous data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The authors thank the support from the National Science Foundation under Grant No. CMMI-2005 661 and No. CMMI-2005 665 and Ford-Northwestern Alliance Project and Intersection Science Fellowship at Northwestern University.

## References

- [1] D. Braha, N. Suh, S. Eppinger, M. Caramanis, and D. Frey, "Complex engineered systems," in *Unifying Themes in Complex Systems*, pp. 227–274, Springer, New York, NY, USA, 2006.
- [2] C. S. Holling, "Understanding the complexity of economic, ecological, and social systems," *Ecosystems*, vol. 4, no. 5, pp. 390–405, 2001.
- [3] C. Hoyle, W. Chen, N. Wang, and F. S. Koppelman, "Integrated Bayesian hierarchical choice modeling to capture heterogeneous consumer preferences in engineering design," *Journal of Mechanical Design*, vol. 132, no. 12, Article ID 121010, 2010.
- [4] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
- [5] Z. Tao and H. Zhang, "Partnering strategies of organizational networks in complex environment of disaster in the centralized political context," *Complexity*, vol. 2020, Article ID 9687390, 13 pages, 2020.
- [6] Z. Ball and K. Lewis, "Observing network characteristics in mass collaboration design projects," *Design Science*, vol. 4, 2018.
- [7] J. Liu, Z. Chong, and S. Lu, "The evolution and determinants of interorganizational coinvention networks in new energy vehicles: evidence from shenzhen, China," *Complexity*, vol. 2021, Article ID 6665945, 12 pages, 2021.
- [8] Q. Wu and K. Cormican, "Shared leadership and team creativity: a social network analysis in engineering design teams," *Journal of Technology Management & Innovation*, vol. 11, no. 2, pp. 2–12, 2016.
- [9] Z. Sha, A. M. Chaudhari, and J. H. Panchal, "Modeling participation behaviors in design crowdsourcing using a bipartite network-based approach," *Journal of Computing and Information Science in Engineering*, vol. 19, no. 3, 2019.
- [10] S. Li, L. Tong, J. Xing, and Y. Zhou, "The market for electric vehicles: indirect network effects and policy design," *Journal of the Association of Environmental and Resource Economists*, vol. 4, no. 1, pp. 89–133, 2017.
- [11] V. Breschi, M. Tanelli, C. Ravazzi, S. Strada, and F. Dabbene, "Social network analysis of electric vehicles adoption: a data-based approach," in *Proceedings of the 2020 IEEE International Conference on Human-Machine Systems (ICHMS)*, pp. 1–4, IEEE, Rome, Italy, September 2020.
- [12] D. Wu, D. Schaefer, and D. W. Rosen, "Cloud-based design and manufacturing systems: a social network analysis," *Design for Harmonies*, vol. 7, pp. 19–22, 2013.
- [13] Z. Sha, Y. Huang, J. S. Fu et al., "A network-based approach to modeling and predicting product coconsideration relations," *Complexity*, vol. 2018, Article ID 2753638, 14 pages, 2018.
- [14] M. Wang and W. Chen, "A data-driven network analysis approach to predicting customer choice sets for choice modeling in engineering design," *Journal of Mechanical Design*, vol. 137, no. 7, 2015.
- [15] M. Wang, W. Chen, Y. Huang, N. S. Contractor, and Y. Fu, "Modeling customer preferences using multidimensional network analysis in engineering design," *Design Science*, vol. 2, 2016.
- [16] M. Wang, Z. Sha, Y. Huang, N. Contractor, Y. Fu, and W. Chen, "Forecasting technological impacts on customers' co-consideration behaviors: a data-driven network analysis approach," in *Proceedings of the ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, ASME, Charlotte, NC, USA, December 2016.
- [17] J. S. Fu, Z. Sha, Y. Huang, M. Wang, Y. Fu, and W. Chen, "Modeling customer choice preferences in engineering design using bipartite network analysis," in *Proceedings of the ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, AMSE, Cleveland, OH, USA, November 2017.
- [18] M. Wang, Z. Sha, Y. Huang, N. Contractor, Y. Fu, and W. Chen, "Predicting product co-consideration and market competitions for technology-driven product design: a network-based approach," *Design Science*, vol. 4, 2018.
- [19] Y. Bi, J. Xie, Z. Sha, M. Wang, Y. Fu, and W. Chen, "Modeling spatiotemporal heterogeneity of customer preferences in engineering design," in *Proceedings of the ASME 2018 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, November 2018.
- [20] F. Ahmed, Y. Cui, Y. Fu, and W. Chen, "A graph neural network approach for product relationship prediction," in *Proceedings of the ASME 2021 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, August 2021.
- [21] M. Wang, W. Chen, Y. Huang, N. S. Contractor, and Y. Fu, "A multidimensional network approach for modeling customer-product relations in engineering design," in *Proceedings of the ASME 2015 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, January 2015.
- [22] P. N. Krivitsky, "Exponential-family random graph models for valued networks," *Electronic journal of statistics*, vol. 6, pp. 1100–1128, 2012.
- [23] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, "An introduction to exponential random graph (p\*) models for social networks," *Social Networks*, vol. 29, no. 2, pp. 173–191, 2007.
- [24] S. Wasserman and P. Pattison, "Logit models and logistic regressions for social networks: I. an introduction to Markov graphs and p," *Psychometrika*, vol. 61, no. 3, pp. 401–425, 1996.
- [25] J. Xie, Y. Bi, Z. Sha et al., "Data-Driven dynamic network modeling for analyzing the evolution of product competitions," *Mechanical Design*, vol. 142, no. 3, Article ID 031112, 2019.
- [26] M. S. Handcock, D. R. Hunter, C. T. Butts, S. M. Goodreau, and M. Morris, "statnet: software tools for the representation, visualization, analysis and simulation of network data," *Journal of Statistical Software*, vol. 24, no. 1, pp. 1548–7660, 2008.
- [27] D. R. Hunter, M. S. Handcock, C. T. Butts, S. M. Goodreau, and M. Morris, "ergm: a package to fit, simulate and diagnose exponential-family models for networks," *Journal of Statistical Software*, vol. 24, no. 3, Article ID 54860, 2008.

- [28] J. Van Der Pol, "Introduction to network modeling using Exponential Random Graph models (ERGM). working paper or preprint," *Computational Economics*, vol. 54, no. 1, 2017.
- [29] P. N. Krivitsky and C. T. Butts, *Modeling Valued Networks with Statnet*, p. 2013, The Statnet Development Team, 2013.
- [30] T. A. Scott, "Analyzing policy networks using valued exponential random graph models: do government-sponsored collaborative groups enhance organizational networks?" *Policy Studies Journal*, vol. 44, no. 2, pp. 215–244, 2016.
- [31] A. Pilny and Y. Atouba, "Modeling valued organizational communication networks using exponential random graph models," *Management Communication Quarterly*, vol. 32, no. 2, pp. 250–264, 2018.
- [32] M. J. Silk, N. L. Weber, L. C. Steward et al., "Contact networks structured by sex underpin sex-specific epidemiology of infection," *Ecology Letters*, vol. 21, no. 2, pp. 309–318, 2018.
- [33] M. Windzio, "The network of global migration 1990-2013," *Social Networks*, vol. 53, pp. 20–29, 2018.
- [34] A. De Montis, M. Barthélemy, A. Chessa, and A. Vespignani, "The structure of interurban traffic: a weighted network analysis," *Environment and Planning B: Planning and Design*, vol. 34, no. 5, pp. 905–924, 2007.
- [35] G. Robins, "A tutorial on methods for the modeling and analysis of social network data," *Journal of Mathematical Psychology*, vol. 57, no. 6, pp. 261–274, 2013.
- [36] M. S. Handcock, D. R. Hunter, C. T. Butts et al., "Statnet: software tools for the statistical analysis of network data," *Journal of Statistical Software*, vol. 24, no. 1, pp. 1548–7660, 2019.
- [37] T. A. Snijders, "Markov chain Monte Carlo estimation of exponential random graph models," *Journal of Social Structure*, vol. 3, no. 2, pp. 1–40, 2002.
- [38] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [39] K. P. Burnham and D. R. Anderson, "Multimodel inference," *Sociological Methods & Research*, vol. 33, no. 2, pp. 261–304, 2004.
- [40] A. D. Stivala, H. C. Gallagher, D. A. Rolls, P. Wang, and G. L. Robins, "Using sampled network data with the auto-logistic actor attribute model," 2020, <https://arxiv.org/abs/2002.00849>.
- [41] C. R. Shalizi and A. Rinaldo, "Consistency under sampling of exponential random graph models," *Annals of Statistics*, vol. 41, no. 2, pp. 508–535, 2013.
- [42] R. He and T. Zheng, "Estimation of exponential random graph models for large social networks via graph limits," in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 248–255, IEEE, Niagara Falls, ON, Canada, August 2013.
- [43] W. An, "Fitting ERGMs on big networks," *Social Science Research*, vol. 59, pp. 107–119, 2016.
- [44] A. Stivala, G. Robins, and A. Lomi, "Exponential random graph model parameter estimation for very large directed networks," *PLoS One*, vol. 15, no. 1, Article ID e0227804, 2020.
- [45] F. Yin and C. T. Butts, "Kernel-based approximate Bayesian inference for exponential family random graph models," 2020, <https://arxiv.org/abs/2004.08064>.