Networks and Spatial Economics
https://doi.org/10.1007/s11067-021-09526-9

Modeling Multi-Year Customers' Considerations and Choices in China's Auto Market Using Two-Stage Bipartite Network Analysis

Youyi Bi[1] · Yunjian Qiu[2] · Zhenghui Sha[3] · Mingxian Wang[4] · Yan Fu[4] · Noshir Contractor[5] · Wei Chen[6]

Accepted: 1 March 2021/Published online: 26 March 2021

Abstract

Choice modeling is important in transportation planning, marketing and engineering design, as it can quantify the influence of product attributes and customer demographics on customers' choice behaviors. Consumer studies suggest that customers' choice-making process often consists of two different stages: customers first consider subsets of available products on the market, and then make the final choice from the subsets. As existing preference modeling is mostly focused on the choice stage, there is a need to develop methods for understanding customer preferences at both stages, and investigate how customer preferences change from "consideration" to "choice", and whether such changes will be consistent over time. In this paper, we study customers' consideration and purchase behaviors in China's auto market using multi-year survey datasets. We demonstrate how descriptive network analysis and analytic network models (bipartite Exponential Random Graph Model (ERGM)) capture the change of customers' preferences from the consideration stage to the choice stage in multiple consecutive years. Our results show that factors such as fuel consumption per unit power, car make origin, and place of production influence customers' considerations and final purchase decisions in different ways, and this difference between consideration and purchase is consistent over time. The main contribution of this study is that we validate the two-stage network-based modeling approach and its utility in preference elicitation using multiple-year dataset, which sheds lights on understanding the trend of customers' consideration and choice behaviors across years. Our study also contributes to a refined interpretation of the ERGM results with categorization of continuous variables into ranges, which shows that customer choice decisions may be more qualitatively influenced by product attributes rather than quantitatively. Our approach is generic and thus can be applied to solving broader choice modeling problems, such as the transportation mode selection and the adoption of clean technology (e.g., electric vehicles).

✉ Wei Chen
weichen@northwestern.edu

Extended author information available on the last page of the article

Keywords Choice model · Network analysis · ERGM · Consideration · Customer preference · Auto market

1 Introduction

Quantitative choice models are critical for demand estimation in transportation systems (Ben-Akiva and Lerman 1985; Fujii and Kitamura 2004; Tanaka et al. 2014), marketing (Kamakura and Russell 1989; Green and Srinivasan 1990) and engineering design (Chen et al. 2013; Sha et al. 2017; Bi et al. 2018). Modeling individual choices is inherently difficult due to the challenges in modeling various forms of heterogeneity in human behaviors (Shocker et al. 1991; Amador et al. 2008), such as the variance between customers' multi-phased decision processes (Shao 2007). Recent consumer studies suggest that customers' decision-making process often consists of two different stages: customers first form a consideration set from the available products in the market, then make the final choice from that consideration set (Hauser and Wernerfelt 1990; Hauser et al. 2010). Such a two-stage consider-then-choose process becomes more common when customers are facing a large number of alternatives (Hauser et al. 2010; Wasi et al. 2012), such as purchasing vehicles. As Hauser et al. (2009) stated, "*if customers do not consider your product, they can't choose it.*" It is therefore important to understand both customers' consideration preferences and choice preferences in order to accurately model the two-stage decision-making process.

Existing studies of customer preference modeling mostly focus on the final choice decisions (Hoyle et al. 2010; He et al. 2012, 2014). Studies that have taken both consideration and choice stages into account are often limited in modeling customers' compensatory behaviors in both stages. In consumer studies (Wright 1975), compensatory behavior means a customer arrives at a choice by averaging or adding the positive attributes of a product and subtracting the negatives (e.g., choosing a car by making tradeoffs among price and performance), and non-compensatory behavior means a customer eliminates (chooses) an alternative from the choice set because of a negative (positive) evaluation of any one attribute which cannot be offset by another attribute (e.g., choosing a laptop model just because it has a special feature that other models do not). For example, Liu and Arora (2011) developed a method to construct efficient experimental designs for customer choices using a two-stage, consider-then-choose model that involves a non-compensatory screening process at the first stage and a compensatory choice process at the second stage. Ross Morrow et al. (2014) proposed a solution to design optimization problems when demand is modeled with consider-then-choose models in which screening rules are based on conjunctive rules (i.e., an alternative must have a minimum value across several criteria in order to be considered). Even if various forms of rules and strategies have been generalized for capturing customer preferences in the consideration (non-compensatory) stage, there is a lack of quantitative approaches to investigate customers' compensatory behaviors in the stages of consideration and choice jointly. The **main objective of this research** is to study customers' consideration and choice behaviors as distinct but integrated activities using network-based approaches, and investigate how customer preferences change from "consideration" to "choice" based on data collected across multiple years.

In the past decade, disaggregate quantitative models, such as Discrete Choice Analysis (DCA) (Train 1986, 2009; Nielsen et al. 2002; Lam and Huang 2003; Amador et al. 2008) and conjoint analysis (Tovares et al. 2013), have been studied extensively by the choice modeling community. These random-utility-theory based methods (Manski 1977) are better in capturing the uncertainty of individual choices compared to aggregate choice models, such as Multiple Discriminant Analysis (Johnson 2011), Factor Analysis (Gorsuch 1983), and Multi-dimensional Scaling (Green 1970)), which simply divide customers into groups with similar demographic profiles and neglect the individual heterogeneities in decision-making. However, DCA is limited by not adequately dealing with the dependency of alternatives, the social influence among individuals, and the collinearity of attributes (Wang et al. 2016a). To overcome these limitations, recent studies in choice modeling have explored the method of using social network attributes in traditional DCA (Rasouli and Timmermans 2016) or leveraging network analysis to model customer preferences (Wang et al. 2015b, 2016a, 2018). Network-based approaches analyze and model complex relationships based on a networked graph, where nodes represent individual members and links represent relationships between members (Saberi et al. 2018). By considering the influence of both exogenous effects (e.g., nodal attributes) and endogenous effects (e.g., network structures) on the formation of links in a network, network-based approaches relax the independence assumptions of individuals or alternatives in DCA and are capable of modeling complex social behavior and cross-level interactions (Robins et al. 2007; Wang et al. 2016a).

In our preliminary work (Fu et al. 2017), we presented a two-stage network-based modeling approach to study customers' consideration and choice behaviors. The key idea of the proposed approach is that we consider customers and products as two types of nodes in a bipartite network and model customers' choice behaviors (i.e., the formation of links) in two distinct but integrated stages: the consideration behaviors in Stage 1 and the purchase behaviors in Stage 2. This approach allows us to quantify the influence of both exogenous effects (e.g., customer and product attributes) and endogenous effects (e.g., market distribution) on customers' choice-making decisions, and investigate how customer preferences change from "consideration" to "choice". The results obtained from a single-year analysis show that the factors influencing customers' consideration decisions are different from those influencing their final purchase decisions. In addition, both exogenous factors (i.e., car attributes and customer demographics) and the endogenous network structural factors (i.e., existing vehicle competition relationships) significantly influence both customers' consideration and choice decisions. The results also indicate that the two-stage network-based approach outperforms logistic regression and the one-stage network-based approach (i.e., assuming every customer makes a choice directly from all available products in the market) in modeling complex and interdependent customer preferences. However, these research findings are based on the data from a single-year survey. A more thorough investigation using multi-year dataset is necessary to test whether these findings are consistent over time.

Since customer preferences may evolve over time (Min and Han 2005; Bi et al. 2017, 2018), we are also interested in better understanding the evolving patterns of consideration-then-choice preference over time. For example, fuel-efficient cars became more preferred after the energy crisis in the 1970s (Turrentine and Kurani 2007). Around the 2000s, research on the luxury vehicle market found that the symbolic values of luxury vehicles dominated over their functional characteristics (Dubois et al.

2001). Dubois and Paternault (1995) stated, "*More than other products, luxury items are bought for what they mean, beyond what they are*". However, more recent research (Stylidis et al. 2016) reveals that the understanding of perceived quality requirements and attributes has become more important for customers who purchase luxury vehicles, suggesting car companies should produce luxury vehicles more comparable to the premium segment vehicles (here premium is a segment between economy and luxury) with respect to functionality and quality. These examples indicate that customer preferences could possibly change over time, and more empirical evidence is needed to prove that. Understanding the change of customer preferences is important for estimating dynamic demand.

In this paper, we examine the two-stage network-based modeling approach using multiple-year survey data from China's auto market as an example, and investigate the temporal change of customers' consideration and choice behaviors over the years. The main contribution of this work is to further validate the use of the two-stage network-based modeling approach in capturing customers' preferences over time, and provide more empirical evidence to better understand how customer preferences move from "consideration" to "choice". In particular, we are interested in answering the following research questions:

- *How does the structure of the customer-product networks change over time?*
- *How do the factors influencing customers' consideration of products differ from those influencing their final purchase decisions? Is the finding consistent across different years?*

The remainder of this paper is structured as follows. Section 2 provides an overview of the use of network analysis in choice modeling and the two-stage network-based modeling approach. Section 3 begins by introducing the dataset utilized in this research as well as the results from descriptive analysis on the explanatory variables and network structures. Then we present the ERGM results and discuss the temporal change of the estimated coefficients. Given that customers preferences are influenced by broader categories rather than specific values of variables, we next present results based on a refined ERGM that bins continuous data into categories. Hence, this work also contributes to refining the implementation and interpretation of the two-stage network-based approach. Section 4 summarizes the modeling results and their implications with recommendations for future work.

2 Technical Background

In this section, the technical background of network analysis in choice modeling is introduced, and the two-stage network-based modeling approach is reviewed.

2.1 Network Analysis in Choice Modeling

As a system exhibiting dynamic, uncertain, and emerging behaviors, customer-product relations can be viewed as a complex socio-technical system and analyzed using network theory and techniques. Our earlier work (Wang et al. 2015a, 2016a) demonstrated that customer preferences can be modeled in the form of customer-product relations in a multidimensional network. In this network, nodes represent customers and products in multiple layers of the network, and links represent the relations among these nodes, such as the social interactions between customers (in customer layer), the product association (i.e., similarity) between products (in product layer), and the consideration or purchase decisions (between customer and product layers).

Network-based approach offers both descriptive network analysis and analytical network models. Descriptive network analysis is powerful in visualizing the topological characteristics identified in customer-product networks (Wang et al. 2015a). Descriptive network metrics (e.g., degree, density, closeness, etc.) provide an intuitive understanding on the properties of nodes and links in a network (Ding et al. 2019). In addition to the descriptive capability, recent advance in Exponential Random Graph Models (ERGMs) provides a general and versatile statistical inference framework for analytical network modeling. ERGM takes an observed network, $\boldsymbol{y}$, as a specific instance from a set of possible random networks, $\boldsymbol{Y}$, following the distribution in Eq. (1) (Robins et al. 2007; Broekel and Bednarz 2018).

$$Pr(\boldsymbol{Y}=\boldsymbol{y}) = \frac{\exp\left\{\boldsymbol{\theta}'\boldsymbol{g}(\boldsymbol{y})\right\}}{\kappa(\boldsymbol{\theta})}, \tag{1}$$

where $\boldsymbol{\theta}$ is a vector of model parameters ($\boldsymbol{\theta}'$ is the transpose of $\boldsymbol{\theta}$), $\boldsymbol{g}(\boldsymbol{y})$ is a vector of the network statistics, and $\kappa(\boldsymbol{\theta})$ is a normalizing factor to ensure Eq. (1) is a proper probability distribution. Equation (1) suggests that the probability of observing a particular network is proportional to the exponent of a weighted combination of network statistics: one statistic $g(y)$ is more likely to occur if the corresponding θ is positive. The estimated coefficients of ERGMs $\boldsymbol{\theta}$ can quantify the influence of the network statistics $\boldsymbol{g}(\boldsymbol{y})$ (i.e., explanatory variables such as nodal attributes, network structures) on the formation of links in a network. For example, if the estimated coefficient of a star term (a network structure) is significantly positive in an ERGM, then the nodes tend to connect with those nodes that already have a large number of links in the observed network. In ERGMs, the network itself is a random variable and the probability is evaluated on the entire network instead of a link. Markov Chain Monte Carlo (MCMC) simulation is often used to estimate the parameter $\boldsymbol{\theta}$ that maximize the likelihood of observed network structures at the aggregate level when fitting an ERGM.

In our prior work, ERGMs have been used to study customers' consideration behaviors using the unidimensional network at the aggregated market level (Sha et al. 2017) and multidimensional network at the disaggregated customer level (Wang et al. 2016a), respectively. The estimated model was also used to forecast the impact of technological changes (e.g., turbo engine) on market competitions (Wang et al. 2016c, 2018), and predict product co-consideration relations (Sha et al. 2018). Compared to the random utility-based customer preference models such as Discrete Choice Analysis (DCA) (Train 2009), our results have illustrated that network-based models can more effectively handle the *interdependency* among products, the *social influence* on customers, and the *collinearity* of attributes (Wang et al. 2016a).

Considering these advantages, ERGM is chosen as the statistical inference framework of the two-stage network-based modeling approach reviewed in the following subsection.

2.2 The Two-Stage Network-Based Modeling Approach (Fu et al. 2017)

Two-stage choice model assumes each customer considers a subset of available products on the market first, and then makes the final choice from the subset. Figure 1 illustrates the bipartite networks corresponding to the two-stage model. Unlike one-mode networks consisting of a single layer of nodes (e.g., social influence network (He et al. 2014) and co-consideration network (Wang et al. 2016a, 2016b)) or multi-dimensional networks consisting of multiple types of nodes with relations among them, a bipartite network (Zhao et al. 2020) is comprised of links (e.g., consideration and purchase) connecting two layers without relations among the nodes within the same layer (e.g., the customer layer or the product layer). In Stage 1, consideration links (dash lines) represent customers' considerations among all possible product alternatives. In Stage 2, purchase links (solid lines) indicate the final purchase decisions among all products being considered in Stage 1. For example, customer C1 considered products P1 and P2 in Stage 1, and purchased product P1 in Stage 2. To train the two-stage model, both the consideration sets and the final choices data must be available. Once a full two-stage model is established, both the consideration set and the final choice can be predicted using the Stage 1 and Stage 2 models sequentially.

Bipartite ERGM is employed in the two-stage network-based modeling to quantify the influence of both exogenous effects and endogenous effects on the formation of consideration and choice relations. One example of endogenous effect is the "geometrically weighted degree distribution" that measures the evenness of the distribution of nodal degrees in a network. Bipartite ERGM is a type of ERGMs specifically for modeling bipartite network structures (Wang 2013; Wang et al. 2013a, 2013b), following the same model structure as defined in Eq. (1) where $\boldsymbol{g}(\boldsymbol{y})$ captures network statistics within the bipartite affiliation network. Eq. (2) illustrates the connection between the models in two stages, where $\boldsymbol{y}_1$ is the observed network of Stage 1 (consideration), $\boldsymbol{y}_2$ is the observed network of Stage 2 (purchase), and $\boldsymbol{y}_1^-$ is the complementary set of $\boldsymbol{y}_1$ (i.e., an artificial network that assumes customers consider all products that were not actually considered). Eq. (2) indicates that the probability of the purchase links in Stage 2 is conditional on the consideration network in Stage 1, and the probability of the purchase links between customers and products outside

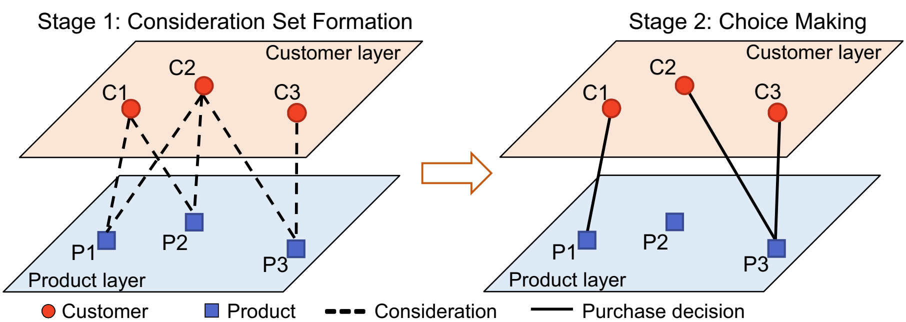

Fig. 1 Bipartite network for two-stage consider-then-choose modeling approach

their consideration sets is zero. For more details on the implementation of this approach in the R environment, please refer to (Hunter et al. 2008; Morris et al. 2008; Fu et al. 2017).

$$Pr(\boldsymbol{Y} = \boldsymbol{y_2}|\boldsymbol{y_1}) = \frac{\exp\left\{\boldsymbol{\theta}'\boldsymbol{g}(\boldsymbol{y}_2)\right\}}{\kappa(\boldsymbol{\theta})}, Pr\left(\boldsymbol{Y} = \boldsymbol{y_2}|\boldsymbol{y_1^-}\right) = 0 \tag{2}$$

3 Dataset and Results

3.1 Dataset and Descriptive Analysis of Modeling Attributes

In this study, we employ the data from a recognized, reputable survey representing China's auto market (Fu et al. 2017) collected by a global market research firm. This survey data consists of the response and purchase history of about 50,000 to 70,000 new car buyers per year and covers about 400 vehicle models in China's market from 2013 to 2015. Respondents were asked to list the cars they purchased with up to two alternative cars they considered before the purchase. The vehicle's attributes, such as engine power and fuel consumption, are reported by customers in the survey and verified by the data company. This survey data also consists of respondents' demographics. For demonstration purposes, we choose a particular market segment for analysis. This enables our analysis to focus on a self-contained choice set in which preferences tend to be consistent. As the most popular segment, the compact sedan chosen for this study accounts for about one-third of the respondents' choices in all 3 years as shown in the survey. In the survey, not all respondents reported the other cars they considered in addition to the cars they purchased. To ensure each customer in our study made tradeoffs between different cars before he or she made the final purchase decision, we only considered respondents who listed at least two car models in their consideration sets and selected a compact sedan as their final purchase. However, the considered car models do not have to be sedans. This filtering gives us 18,056, 22,196, and 25,896 customers from 2013 to 2015, respectively.

Based on our previous research of customer preferences in vehicle considerations and choices (Fu et al. 2017), we considered five key vehicle attributes, *price*, *fuel consumption*, *power*, *imported, make origin*, and one customer attribute, *Tier 1,* for this study. *Price* is a continuous variable, referred as the average price paid to dealers of a vehicle. *Fuel consumption* is a continuous variable measuring the liters of fuel that a vehicle consumes per 100 km. *Power* is a continuous variable, representing the average engine power of a vehicle in brake horsepower (BHP). *Imported* is a binary variable, indicating whether a vehicle is domestically produced (*Imported* = 0) or imported (*Imported* = 1). *Make origin* is a categorical variable showing the origin of the car brand. Although some vehicles are produced in China (i.e., *Imported* = 0), their make origin could be Korea, Japan, US, or Europe. *Tier 1* is a binary variable describing whether a customer comes from tier-one cities (*Tier 1* = 1) or not (*Tier 1* = 0). In this study, tier-one cities consist of Beijing, Shanghai, and Guangzhou, which are the largest and most prosperous cities in China.

Table 1 provides a descriptive analysis of the selected vehicle and customer attributes of the purchased car models. Here N denotes either the number of car models in the compact sedan segment or the number of customers after filtering. It can be seen that the mean value of *price* and *power* are consistent throughout the 3 years, but the average fuel consumption and percentage of imported cars decreased. This may indicate that the compact sedans with lower fuel consumption and those produced domestically were increasingly popular with Chinese customers. In addition, European cars held the largest share of the compact sedans with foreign brands sold in China, although their proportion decreased from 2013 to 2015. Japanese cars became more popular, and the share of Korean cars decreased. Another interesting finding is that the percentage of customers from tier-one cities in all respondents decreased. This may imply that more customers from economically developed areas (tier-one cities) purchased cars from other segments, and compact sedans were spreading in those less economically developed areas in China from 2013 to 2015.

3.2 Descriptive Analysis on Network Structures

Network visualizations and descriptive metrics can provide us an intuitive understanding of the network structures at different decision-making stages or time periods. Figure 2 presents six representative local network structures in the consideration stage and choice stage. They are called local because these plots are based on 20 randomly chosen customers who all considered a Ford Changan Focus (2012 MY) between 2013 and 2015. In each plot, a red circle represents a customer with his/her ID number, a blue square represents a car with its model name, and a black line denotes the consideration or purchase relationship between them. As mentioned before, links only exist between different types of nodes in a bipartite network. Each customer may consider two or three car models in Stage 1, and purchases one car in Stage 2. Each car can be considered or purchased by multiple customers. Some car models (e.g., Skoda SVW Octavia in Fig. 2 (a)) own more links than others, that is, they have a larger *degree* (i.e., number of connections) in the network, indicating they are more popular for customers to consider or purchase. Table 2 provides the representative metrics of the two-stage networks from 2013 to 2015, such as the number of nodes/links, the average degree (the average number of links that a node has) and the network density (the proportion of present links from all possible links in the network). For consistent comparisons, 5000 customers each year are randomly chosen from the filtered datasets, in which the percentage of customers from tier-one cities (approximately 21% in 2013) is consistent across years to reduce the potential influence from the systematic differences of the samples of 3 years. We observe that the ratio of the number of car nodes in Stage 1 to that in Stage 2 and the network density in both stages are consistent through the 3 years, indicating that the randomly chosen datasets are comparable to each other.

As shown in Fig. 2, although Ford Changan Focus was considered by 20 customers, eventually only a few customers chose to purchase this model. This implies that customers do not always purchase those cars frequently considered and they may use different judging criteria when making decisions in consideration and purchase stages. One can observe that the consideration networks in 2014 and 2015 are denser than that in 2013. This implies that more car models enter China's auto market as time goes by, and customers' considerations are becoming more diversified. This implication is

Table 1 Descriptive statistics of the selected vehicle and customer attributes of purchased car models from 2013 to 2015

Continuous Variables							
		N	Mean	St. Dev	Min	Max	
Price (RMB)	2013	84	113,933.90	50,292.02	48,852.45	335,716.90	
	2014	96	116,480.20	52,052.65	50,048.08	400,980.00	
	2015	99	110,119.70	48,096.41	50,214.75	364,250.00	
Fuel Consumption (liter/100 km)	2013	84	8.45	0.87	7.06	11.88	
	2014	96	8.30	0.75	6.97	11.11	
	2015	99	8.02	0.71	6.88	10.70	
Power (BHP)	2013	84	123.54	19.82	80.39	205.00	
	2014	96	122.89	17.77	90.00	208.00	
	2015	99	124.20	19.05	94.56	212.00	
Binary Variables							
		N	Min	Frequency of Min	Max	Frequency of Max (%)	
Imported	2013	84	0	76	1	8 (9.50%)	
	2014	96	0	90	1	6 (6.25%)	
	2015	99	0	93	1	6 (6.06%)	
Tier 1	2013	18,056	0	14,227	1	3829 (21.21%)	
	2014	22,196	0	18,603	1	3593 (16.19%)	
	2015	25,896	0	22,733	1	3163 (12.21%)	
Categorical Variable							
		N	China (%)	USA (%)	Europe (%)	Japan (%)	Korea (%)
Make Origin	2013	84	33 (39.29%)	6 (7.14%)	23 (27.38%)	15 (17.86%)	7 (8.33%)
	2014	96	37 (38.54%)	6 (6.25%)	25 (26.04%)	20 (20.83%)	8 (8.33%)
	2015	99	38 (38.38%)	10 (10.10%)	24 (24.24%)	22 (22.22%)	5 (5.05%)

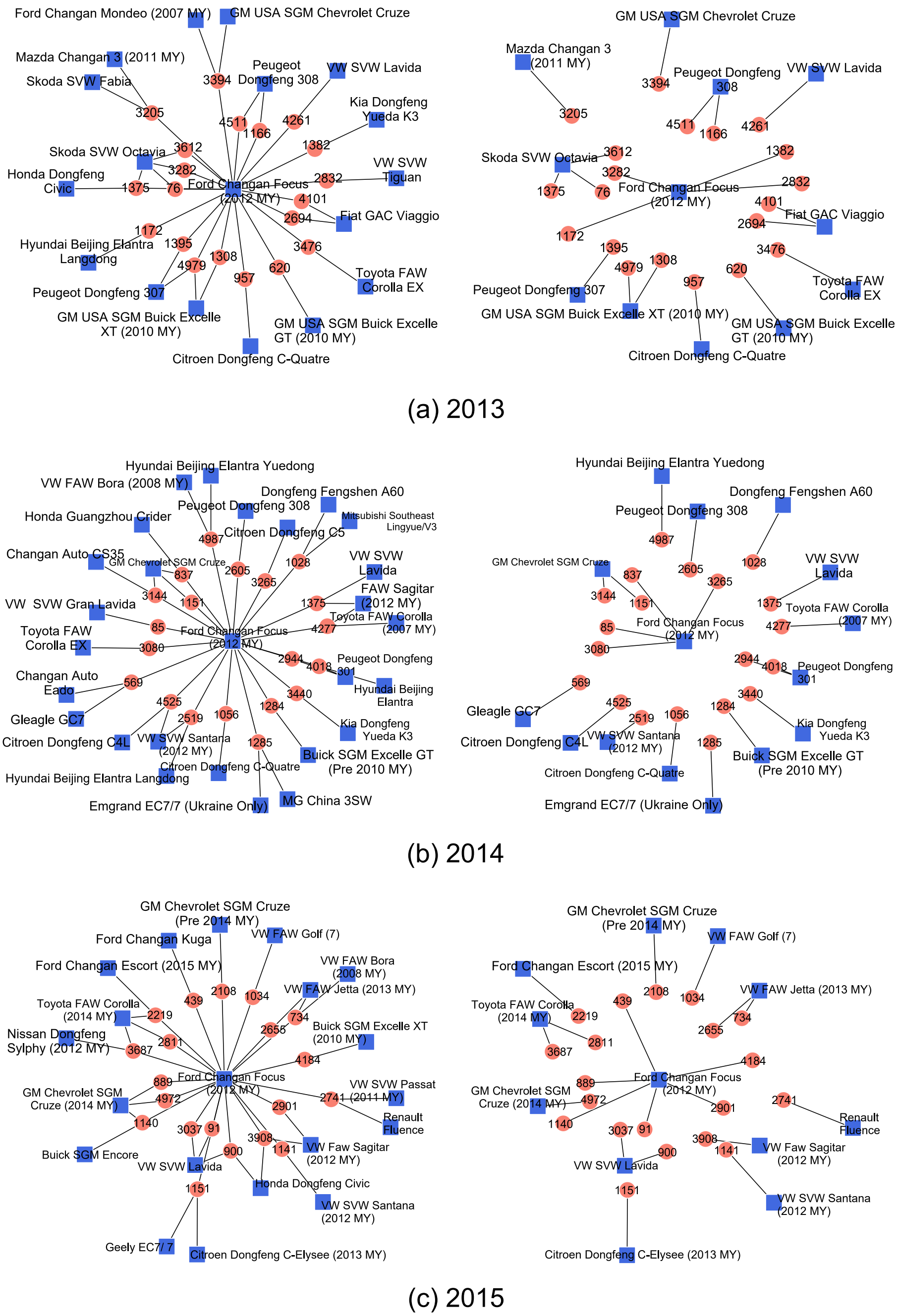

Fig. 2 Comparison of representative local network structures, where a red circle represents a customer with his/her ID number and blue squares are car models with names. The left plots in (**a**), (**b**) and (**c**) are networks of Stage 1 – consideration, and the right plots are networks of Stage 2 – purchase

further validated by the change of the number of car nodes and average degrees of customer/car nodes from 2013 to 2015 as shown in Table 2. For example, the number

Table 2 Representative descriptive statistics of the two-stage bipartite networks from 2013 to 2015

Year	Stage	# of nodes		# of links	Avg. degree of nodes		Density	% of customer nodes with 3 degrees
		Customer	Car		Customer	Car		
2013	1	5000	243	11,221	2.244	46.177	8.166E-04	24.4%
	2	5000	83	5000	1.000	60.241	3.871E-04	0.0%
2014	1	5000	256	11,387	2.277	44.480	8.327E-04	27.7%
	2	5000	96	5000	1.000	52.083	3.865E-04	0.0%
2015	1	5000	267	11,248	2.250	42.127	8.254E-04	25.0%
	2	5000	99	5000	1.000	50.510	3.878E-04	0.0%

of car nodes in 2014 (256) and 2015 (267) is larger than in 2013 (243) in Stage 1, and the same observation can be found in Stage 2. The average degree of customer nodes in 2014 (2.277) and 2015 (2.250) is larger than in 2013 (2.244) in Stage 1. The average degree of car nodes in 2014 (44.480) and 2015 (42.127) is smaller than in 2013 (46.177) in Stage 1, and a similar observation can be found in Stage 2. In addition, the percentage of customers with 3 degrees in Stage 1 is higher in 2014 (27.7%) and 2015 (25.0%) than in 2013 (24.4%).

3.3 The Modeling Results

3.3.1 The Modeling Procedures

Compared to descriptive network analysis, analytical network models can quantitatively measure the influence of explanatory variables (e.g., customer demographics, product attributes and network structures) on customers' consideration and purchase decisions. In this study, we adopted similar procedures reported in (Fu et al. 2017) to preprocess data, select modeling variables, and construct the ERGMs for the bipartite consideration and purchase networks. The five key vehicle attributes and one customer attribute presented in Sec. 3.1, and the randomly chosen 5000 customers' data each year from 2013 to 2015 in Sec. 3.2 are employed in our modeling process.

To account for the skewed distribution, *log2* transformation is taken on *price*. Then *price* is normalized using Z-score. In addition, we create two terms in estimating the ERGMs. *Fuel consumption over power* (*FCP*) is a term created by dividing *fuel consumption* by *power* with a unit of L/100 km/BHP. The smaller the *FCP*, the greater the car model's fuel efficiency. *Tier 1 *FCP* is an interaction term between *Tier 1* and *FCP*, describing whether customers from tier-one cities are more likely to consider or purchase cars with better fuel efficiency. Besides these six exogenous variables (i.e., nodal attribute effects, including *price, FCP, power, make origin, imported,* and *tier 1* FCP*), an additional endogenous variable (i.e. network structures) – *geometrically weighted degree distribution (GWDegree)* – is also considered in the ERGMs. In our application context, this network structure variable describes whether the distribution of the degrees of car nodes is even or skewed. A positive coefficient of this variable indicates that most car models have similar numbers of considerations or purchases (an even market), while a negative one indicates that some car models have many more considerations or purchases than others (an uneven market). In the results, we label this network structure variable as *market distribution*.

To construct the bipartite ERGMs for the proposed two-stage model, we first link each customer to the two or three cars he/she considered from the market. We then create a binary variable to indicate whether the considered cars were a customer's final choice, and set the parameter estimates of the cars not being considered for each customer as negative infinity (−inf) in "Statnet" (an R package for statistical network modeling). This operation guarantees the probabilities of customers' purchasing products outside their consideration sets are infinitely close to zero (see Eq. (2)). We then predict the final purchase based on the cars in the consideration set.

3.3.2 The Results of ERGMs

Table 3 provides the estimated parameters ($\boldsymbol{\theta}$) of the two-stage ERGMs (see Eq. (1)) in 2013, 2014, and 2015. We use β to represent an element of the vector $\boldsymbol{\theta}$ in the following paragraphs.

Stage 1 Model: Consideration Only Results from ERGM models[1] show that the endogenous variable, *market distribution*, is significant in 2013 ($\beta = -2.088, p < 0.001$), 2015 ($\beta = -1.202$, $p < 0.001$), but becomes insignificant in 2014 ($\beta = -0.623$, $p > 0.05$). This result suggests that China's auto market might be less skewed in 2014 than in 2013 and 2015, i.e., customers' considerations became more diversified in 2014, which is consistent with the increasing number of car nodes and average degrees of customer nodes as shown in Table 2. The estimated coefficients of exogenous variables are mostly consistent throughout the 3 years. In Stage 1, cars with lower FCP, lower power, and foreign brands are more likely to be considered, although the effect of foreign brands on consideration is weaker in 2015 than in 2013. Domestically produced cars are more likely to be considered by customers in all 3 years, and this effect in 2015 ($\beta = -2.533$, $p < 0.001$) is slightly stronger than in 2013 ($\beta = -2.300$, $p < 0.001$) and 2014 ($\beta = -2.429$, $p < 0.001$). In addition, customers from tier-one cities prefer to consider more fuel-efficient cars in 2013 and 2014. However, the influence of price on customers' considerations has changed between years. Customers prefer to consider cars with lower prices in 2013 ($\beta = -0.530, p < 0.001$) and 2014 ($\beta = -0.465, p < 0.001$) on average, but prefer to consider cars with higher prices in 2015 ($\beta = 0.488, p < 0.001$).[2]

Stage 2 Model: Purchase Conditional on Consideration Results show that the factors influencing customers' consideration of vehicles differ from those influencing their final purchase decisions. Given the consideration sets reported in the survey, the significance of *market distribution* is consistent in Stage 2 throughout the 3 years, implying customers tend to purchase those popular cars in the end. According to the

[1] "Edges" are defined as the number of links in the network, and its coefficient can be considered as the intercept of a regression. However, here the intercept by itself does not necessarily provide meaningful interpretation. To explain the intercept, one needs to set all input categorical variables as baselines and all continuous variables as zeros, which is not physically meaningful in our case. Network researchers commonly interpret the meaning of "Edges" only when it is the single input variable in an ERGM.

[2] We studied an additional dataset with 5000 non-overlapped random samples, and found the new results are quite similar to what we reported here, and the same conclusions hold.

Table 3 Results of the bipartite ERGMs from 2013 to 2015

	2013		2014		2015	
	Stage 1[a]	Stage 2[b]	Stage 1[a]	Stage 2[b]	Stage 1[a]	Stage 2[b]
Edges	−4.160***	−1.151*	−2.374***	0.862.	1.175***	−0.197
Market distribution	−2.088***	−21.00***	−0.623.	−9.170***	−1.202***	−8.265***
Price	−0.530***	−0.681***	−0.465***	−0.331***	0.488***	−0.490***
FCP	−2.709***	0.429*	−4.532***	−0.040	−3.900***	1.631***
Power	−0.013***	0.006*	−0.025***	−0.007**	−0.046***	−0.003
Make origin (US)	2.025***	−0.561***	1.878***	−0.547***	1.474***	−0.338***
Make origin (Europe)	2.183***	−0.096	2.021***	−0.170.	1.673***	0.189*
Make origin (Japan)	0.785***	−0.067	0.806***	−0.355***	0.663***	−0.044
Make origin (Korea)	1.193***	−0.770***	1.407***	−0.599***	1.082***	−0.666***
Import	−2.300***	5.414***	−2.429***	2.568***	−2.533***	2.387***
*Tier1*FCP*	−0.356***	0.203	−0.331*	0.132	−0.127	−0.463
AIC	117,653	–	118,685	–	118,755	–
BIC	117,785	–	118,818	–	118,889	–

AIC and BIC are used to evaluate the goodness of model fitting. The smaller their values, the better fit. AIC and BIC values for Stage 2 models are not available because the parameter estimates for the cars not considered are set to *–inf*, thus, AIC and BIC were not available from ERGM.

*Note. *p < .05; **p < .01; ***p < 0.001. FCP* stands for fuel consumption over power

[a] Stage 1 - Consideration only

[b] Stage 2 - Purchase conditional on consideration

survey, in 2013, 33.18% of car sales were dominated by eight car models (VW Sagitar, Lavida, Bora, Golf, Citroen C-Quatre, Peugeot Dongfeng 308, Buick Excelle and Chevrolet Cruze). Such unevenness in the market is also observed in 2014 and 2015, and is well captured by the endogenous effect *market distribution* in the ERGMs.

Although more fuel-efficient cars are preferred in the consideration stage, customers tend to purchase lower fuel-efficient cars in 2013 ($\beta = 0.429$, $p < 0.05$) and 2015 ($\beta = 1.631$, $p < 0.001$), among their consideration sets. The *FCP* is insignificant for customers' purchase decisions in 2014 ($\beta = -0.040$, $p > 0.1$). One possible explanation is that fuel efficiency plays a more important role in customers' consideration decisions than in their purchase decisions. The cars in consideration sets usually have already met customers' expectations on fuel efficiency and customers may pay more attention to other features (e.g., navigation, Bluetooth stereo, etc.) in the purchase stage. In Stage 1, *power* is significantly negative in all 3 years, but it becomes positive in 2013 ($\beta = 0.006$, $p < 0.001$) and insignificant in 2015 ($\beta = -0.003$, $p > 0.1$) in Stage 2, although it is still significantly negative in Stage 2 of 2014 ($\beta = -0.007$, $p < 0.001$). One interesting observation is that the magnitudes of the coefficients of *power* are quite small in Stage 2, which suggests it may have weaker influence on customers' choices than considerations.

Moreover, most estimated coefficients of *make origin* in Stage 2 become negative or insignificant, although they are all significantly positive in Stage 1. This implies that compared to Chinese brands, all foreign brands seem to be preferred by Chinese customers in the consideration stage from 2013 to 2015. However, American brands and Korean brands are not preferred to Chinese brands in the purchase stage in all 3 years. The only exception is the European brands, which are more preferred as final purchases than Chinese brands in 2015 ($\beta = 0.189$, $p < 0.05$). This may be due to the long-standing high reputation of European brands in China (e.g., Volkswagen is the first widely influencing western auto brand in China).

Similarly, in Stage 2, the coefficients of *import* become positive in all 3 years, while in Stage 1 they are all negative. This does not mean Chinese customers generally prefer to purchase imported cars. Instead, most Chinese customers will not consider imported cars in the consideration stage. Only those customers who initially consider imported cars are more likely to purchase imported cars (if a customer does not consider imported cars, his/her preference for "Import" will not be captured by the ERGM in Stage 2). In the 2013 dataset, the average monthly household income for customers who considered imported cars is 14,651 RMB, which is much higher than the overall average income (10,530 RMB). This observation suggests that the socioeconomic status of customers may influence their preferences for the origin of production of vehicles.

The interactions between *FCP* and tier-one-city customers become insignificant in Stage 2 in all 3 years, although they are significant in 2013 and 2014 in Stage 1. One possible explanation is that tier-one cities in China (Beijing, Shanghai and Guangzhou) have stricter regulations on environmental protection and encourage the sales of cars with lower greenhouse gas emissions. Fuel-efficient cars are more likely to enter the consideration sets of tier-one-city customers due to their lower fuel cost and lighter environmental impact. However, when making the final purchase decisions, other factors (e.g. social influence) or car features (e.g. external styling) may play more important roles than *FCP*.

In summary, cars with lower FCP, lower power, and foreign brands are more likely to be considered but not necessarily purchased by customers. Our results show that cars with lower price, larger FCP, European brands, and those produced overseas are preferred in the second stage.

When comparing the estimated coefficients in Stage 2 from 2013 to 2015, we found that the endogenous variable *market distribution* and exogenous variables such as *price*, *make origin* from the US and Korea, *import* and the interaction between FCP and tier-one-city, are almost consistent between different years. However, the influence of *power* is fluctuating between years. *Power* is significantly positive in 2013 ($\beta = 0.006$, $p < 0.001$), significantly negative in 2014 ($\beta = -0.007$, $p < 0.001$), and insignificant in 2015 ($\beta = -0.003$, $p > 0.1$). We will interpret this result in the next subsection by refining the continuous variables into categorical ones based on distinctive ranges. In addition, *make origin* from Europe is significant in 2015 ($\beta = 0.189$, $p < 0.05$), but insignificant in other 2 years. Similarly, *make origin* from Japan is significant in 2014 ($\beta = -0.355$, $p < 0.001$), but insignificant in other 2 years. These inconsistencies reflect the subtle changes of customers' preferences in a dynamic market environment.

3.4 ERGM Results through Categorization of Continuous Variables

When treating the three explanatory variables as continuous variables in Section 3.3, we find that we are not able to explain the ERGM results completely based on our intuition. This is because, in reality, what matters to customers may not be the exact value, but the range into which a vehicle attribute falls. To overcome this limitation, we convert the continuous variables *price, FCP,* and *power* into categorical variables (see column 1 of Table 4), and estimate their corresponding parameters. The levels of the categories are determined by clustering analysis using the survey data. We adopted the X-means method for clustering analysis as it can automatically calculate the best number of clusters by maximizing the Bayesian Information Criterion (BIC) iteratively (Pelleg et al. 2000). The X-mean method is advantageous compared to the methods that need the number of clusters as an input, such as K-means clustering, since the true number of clusters is not always known (Suryadi and Kim 2019). We ran the X-means clustering procedure each time with one single variable (e.g., price) and one single-year data. Then the range for each category was obtained by averaging the ranges of each

Table 4 ERGM results with treating price, FCP and power as categorical variables

	2013		2014		2015	
	Stage 1[a]	Stage 2[b]	Stage 1[a]	Stage 2[b]	Stage 1[a]	Stage 2[b]
Edges	−10.964***	−15.261	−11.817***	−5.319	−11.800***	−17.285
Market distribution	−0.791*	−10.511***	−0.402	−7.878***	−0.724.	−7.509***
Price (<108 k)	2.660***	8.113	3.325***	0.419	1.295***	9.570
Price (108 k–188 k)	3.444***	7.596	3.428***	−0.046	1.865***	9.215
Price (188 k–366 k)	1.566***	6.071	1.214***	−1.76	0.932***	8.916
FCP (<0.060)	2.418***	−0.945***	3.546***	4.928	2.945***	8.402
FCP (0.060–0.070)	1.386***	−0.137	2.849***	5.481	2.271***	8.713
FCP (0.070–0.090)	1.116***	−0.258	2.274***	5.423	1.724***	9.045
Power (<120)	1.871***	7.653	0.693***	−0.307	3.141***	−0.774
Power (120–147)	1.361***	8.468	0.457***	0.175	2.540***	−0.406
Power (147–208)	0.888***	7.974	0.522***	0.151	1.253***	−1.524**
Make origin (US)	1.239***	−0.548***	1.334***	−0.525***	1.189***	−0.588***
Make origin (Europe)	1.610***	−0.092	1.766***	0.091	1.669***	0.149*
Make origin (Japan)	0.519***	−0.225*	0.700***	−0.308***	0.708***	−0.264***
Make origin (Korea)	0.731***	−1.000***	1.144***	−0.693***	0.871***	−0.753***
Import	−1.980***	4.354***	−1.756***	2.867***	−1.819***	2.122***
*Tier1*FCP*	−0.563***	0.274	−0.548***	0.204	−0.434***	0.049
AIC	112,948	–	115,998	–	116,473	–
BIC	113,152	–	116,203	–	116,678	–

The baseline of categorical price: > 366,000 RMB; The baseline of categorical power: > 208 BHP

The baseline of categorical FCP: > 0.090 L/100 km/BHP

AIC and BIC are used to evaluate the goodness of model fitting. The smaller their values, the better fit. AIC and BIC values for Stage 2 models are not available because the parameter estimates for the cars not considered are set to *–inf*, thus, AIC and BIC were not available from ERGM.

Note. $*p < .05$; $**p < .01$; $***p < 0.001$. *FCP* stands for fuel consumption over power

[a] Stage 1 - Consideration only

[b] Stage 2 - Purchase conditional on consideration

cluster over 3 years. The baseline for each categorical variable is provided in the footnotes under Table 4. For example, the car models with prices larger than 366,000 RMB is set as the baseline for categorical price. In Stage 1 of 2013, customers prefer to consider cars of 108,000–188,000 RMB ($\beta = 3.444$, $p < 0.001$) compared to the cars of more than 366,000 RMB (i.e., baseline).

By comparing the magnitudes of the estimated coefficients of categorical *price*, we found that in Stage 1, customers prefer to consider less expensive cars, but not too inexpensive. Cars priced from 108,000–188,000 RMB seem to be most attractive to customers in Stage 1 for all 3 years ($\beta = 3.444$, 3.428, 1.865 in 2013–2015), whereas inexpensive cars priced lower than 108,000 RMB seem to be the second most attractive option ($\beta = 2.660$, 3.325, 1.295) in 2013–2015. In Stage 2, the estimated coefficients of categorical price are no longer significant in all 3 years. One possible explanation is that in Stage 1, customers tend to form consideration sets based on their targeted price range or targeted segmentation (e.g., compact sedans with average price of 115,000 RMB). However, in Stage 2, customers are more likely to make final choices from cars in similar price categories, which means price is no longer critical in the choice stage.

Similarly, in Stage 1, customers prefer to consider cars with better FCP (i.e., lower FCP). Cars with FCP smaller than 0.060 L/100 km/BHP seems to be most attractive to customers in Stages 1 for all 3 years, but not in Stage 2. In Stage 1, customers prefer cars with lower power on average. Cars with power less than 120 BHP are the most attractive to customers in Stages 1 for all 3 years, but not in Stage 2. In addition, cars with power more than 208 BHP are more preferred than the those with power of 147–208 BHP ($\beta = -1.524, p < 0.01$) in 2015 of Stage 2. In summary, the results in Table 4 suggest that customers prefer to consider cars with a lower price, power, and FCP only within certain ranges.

As a complement to the results in Table 3, results from categorization can help us probe into customers' preferences and the nuances of their decision-making in different stages. The ERGM results with the continuous-variable setting indicate the influence of a vehicle attribute on customers' behaviors on average, and the ERGM results with the categorical-variable setting indicate customers' preferences towards specific ranges of a vehicle attribute. For example, the positive coefficient of *price* in Stage 1 of 2015 (see Table 3) indicates that on average, customers prefer to consider expensive cars, which seems to be counterintuitive. However, by looking at the estimated results in the same column of Table 4, we find that the coefficients of all categorical prices are positive. This suggests that when compared to very expensive cars (i.e., those priced more than 300,000 RMB), customers still prefer to consider the less expensive ones. Therefore, these two types of modeling results jointly indicate that customers are more likely to consider the cars that are neither too expensive nor too inexpensive, and particularly in 2015, a dislike of too-inexpensive cars seems to be stronger than in other years. One piece of supporting evidence is that the least expensive cars (priced lower than 108,000 RMB) are in the second most attractive category in Stage 1 of 2013–2015, but the difference of coefficient magnitudes between Price (<108,000) and Price (188,000-366,000) becomes much smaller in 2015. In addition, the median price of all considered cars (125,177 RMB) is higher than the average price of all considered cars (124,983 RMB) in 2015. However, in 2013 and 2014, the median price of all considered cars is lower than the average purchase price. In summary, the change of *Price* coefficient from negative to positive in Stage 1 of 2015 as shown in Table 3 does not imply that expensive cars are always preferred in the consideration stage. This change may serve as an indicator to reflect the subtle shift in customers' preferences over the years for vehicles in certain ranges of prices. Combining the modeling approach with both continuous variables and categorical variables provides us a more comprehensive understanding of customer preferences.

4 Conclusion

In this paper, we study customers' multi-year consideration and purchase behaviors in China's auto market using the two-stage network-based modeling approach. Compared to our previous work (Fu et al. 2017), the contribution of this study includes obtaining new empirical findings in understanding customer preferences as well as further refining the implementation and interpretation of the two-stage network-based modeling approach. First, we validate the two-stage network-based approach in choice modeling using multi-year data. This allows us to examine whether Fu et al.'s findings from single-year data could still be obtained from multi-years' datasets. It also sheds lights on understanding the trend of customers' consideration and choice behaviors across years. Chinese customers are found to consistently prefer cars with lower fuel consumption over power (FCP), foreign brands and domestic places of production in the consideration stage, but prefer cars with higher FCP, domestic brands and overseas places of production in the purchase stage, from 2013 to 2015. However, their preferences towards price, power and certain brands of cars has changed over time. For example, cars with lower prices are preferred to be considered in all years except 2015. Power is a significant factor in customers' purchase stage in 2013 and 2014, but becomes insignificant in 2015. European brands and Japanese brands are significant in customers' purchase stage in 2014 and 2015, respectively, but insignificant in other years. These inconsistencies reflect the subtle fluctuation of customers' preferences in a dynamic market.

Second, our study contributes to a refined interpretation of the ERGM results with categorization of continuous variables. When we applied the two-stage modeling approach in modeling customers' decisions in some other years, a part of the modeling results cannot be explained intuitively. To overcome this limitation, we categorized some continuous variables in the ERGMs to observe the influence of these variables on customer preferences at distinctive levels. We find that when considering cars within certain ranges of price, customers do not always prefer the least expensive ones. This result implies that customers may care more about the price range a vehicle attribute falls into rather than its exact value, i.e., customers tend to evaluate vehicle attributes more qualitatively than quantitively when considering and purchasing vehicles. By combining the modeling results from both continuous-variable setting and categorical-variable setting, we obtain a more detailed understanding on customer preferences. A counterintuitive result obtained from modeling with the continuous-variable setting may indicate a subtle change of customer preferences for specific ranges of a continuous variable. We expect these insights will enable choice modeling researchers in developing better approaches to elicitate, model and interpret customers' true preferences.

Our study demonstrates the benefit of network-based approaches in choice modeling. We illustrate how descriptive network analysis and analytic network models capture the change of customers' preferences over time, and how the results of these two approaches are consistent with each other. Both approaches reveal that more car models entered China's auto market between 2013 and 2015, and customers' considerations are becoming more diversified. We find both exogenous variables and endogenous variables significantly influence customers' decisions, and network-based approaches can capture the interdependencies between customers' decisions.

In the current study, we focused on building network models using existing data to make observations and draw inferences. Respondents could have considered 5 to 6 cars but only reported 2 or 3 in the survey. The choice set of 2–3 alternatives well reflects the final set of considerations. In reality, customers may have experienced more than two stages in decision-making while narrowing down the consideration set. When the data of multi-stage consideration sets is available, our two-stage modeling framework can be ttagqiz to multi-stage modeling, in which the same principle of forming next stage network conditional on the network in the previous stage still applies.

Although demonstrated by using vehicle design as an example, our approach can be also applied in other data-driven choice modeling problems in transportation engineering and clean technology, such as the adoption of electric vehicles. Furthermore, our approach can be ttagqiz by surrogating the disaggregate customer layer in the bipartite customer-product network with an aggregate customer layer, in which individual customer nodes are aggregated to nodes such as geographic regions or customer groups. This extension will allow us to investigate how regional factors (e.g., rural vs. urban, economic power, and regional demographics) or customer segments (e.g., "middle-class family oriented", "modern hedonist", etc.) affect customers preferences and how this influence changes over time. Due to the lack of social network data, a bipartite network approach is tested for two-stage modeling in this work. A more comprehensive two-stage multidimensional network approach is planned to be investigated in our future work.

Acknowledgments The authors gratefully acknowledge the financial support from NSF-CMMI-1436658 and the Ford-Northwestern Alliance Project.

References

Amador FJ, González RM, de Dios OJ (2008) On confounding preference heterogeneity and income effect in discrete choice models. Networks Spat Econ 8:97–108

Ben-Akiva ME, Lerman SR (1985) Discrete choice analysis: theory and application to travel demand. MIT press

Bi Y, Li S, Wagner D, Reid T (2017) The impact of vehicle silhouettes on perceptions of car environmental friendliness and safety in 2009 and 2016: a comparative study. Des Sci 3

Bi Y, Xie J, Sha Z, et al (2018) Modeling Spatiotemporal Heterogeneity of Customer Preferences in Engineering Design. In: ASME 2018 International design engineering technical conferences & computers and information in engineering conference. Quebec City, 2018
Broekel T, Bednarz M (2018) Disentangling link formation and dissolution in spatial networks: an application of a two-mode STERGM to a project-based R&D network in the German biotechnology industry. Networks Spat Econ 18:677–704
Chen W, Hoyle C, Wassenaar HJ (2013) Decision-based design: integrating consumer preferences into engineering design. Springer Science and Business Media
Ding R, Ujang N, Bin HH et al (2019) Application of complex networks theory in urban traffic network researches. Networks Spat Econ 19:1281–1317
Dubois B, Laurent G, Czellar S, others (2001) Consumer rapport to luxury: analyzing complex and ambivalent attitudes. Groupe HEC Jouy-en-Josas
Dubois B, Paternault C (1995) Understanding the world of international luxury brands: the" dream formula."(special issue: research input into the creative process). J Advert Res 35:69–77
Fu JS, Sha Z, Huang Y et al (2017) 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, vol 2017, Cleveland
Fujii S, Kitamura R (2004) Drivers' mental representation of travel time and departure time choice in uncertain traffic network conditions. Networks Spat Econ 4:243–256
Gorsuch RL (1983) Factor analysis. Lawrence Erlbaum Associates, Hillsdale
Green PE (1970) Multidimensional scaling and related techniques in marketing analysis. Allyn and Bacon
Green PE, Srinivasan V (1990) Conjoint analysis in marketing: new developments with implications for research and practice. J Mark 54:3–19
Hauser JR, Ding M, Gaskin SP (2009) Non-compensatory (and compensatory) models of consideration-set decisions. In: 2009 Sawtooth software conference proceedings, sequin WA, vol 2009, Delray Beach
Hauser JR, Toubia O, Evgeniou T, Befurt R, Dzyabura D (2010) Disjunctions of conjunctions, cognitive simplicity, and consideration sets. J Mark Res 47:485–496
Hauser JR, Wernerfelt B (1990) An evaluation cost model of consideration sets. J Consum Res 16:393–408
He L, Chen W, Hoyle C, Yannou B (2012) Choice modeling for usage context-based design. J Mech Des 134: 31007
He L, Wang M, Chen W, Conzelmann G (2014) Incorporating social impact on new product adoption in choice modeling: a case study in green vehicles. Transp Res Part D Transp Environ 32:421–434
Hoyle C, Chen W, Wang N, Koppelman FS (2010) Integrated Bayesian hierarchical choice modeling to capture heterogeneous consumer preferences in engineering design. J Mech Des 132:121010. https://doi.org/10.1115/1.4002972
Hunter DR, Handcock MS, Butts CT et al (2008) Ergm: a package to fit, simulate and diagnose exponential-family models for networks. J Stat Softw 24:nihpa54860
Johnson R (2011) Multiple discriminant analysis: marketing research applications. Marketing Classics Press
Kamakura WA, Russell GJ (1989) A probabilistic choice model for market segmentation and elasticity structure. J Mark Res 26:379–390
Lam WHK, Huang H-J (2003) Combined activity/travel choice models: time-dependent and dynamic versions. Networks Spat Econ 3:323–347
Liu Q, Arora N (2011) Efficient choice designs for a consider-then-choose model. Mark Sci 30:321–338
Manski CF (1977) The structure of random utility models. Theory Decis 8:229–254
Min S-H, Han I (2005) Detection of the customer time-variant pattern for improving recommender systems. Expert Syst Appl 28:189–199
Morris M, Handcock MS, Hunter DR (2008) Specification of exponential-family random graph models: terms and computational aspects. J Stat Softw 24:1548–7660
Nielsen OA, Daly A, Frederiksen RD (2002) A stochastic route choice model for car travellers in the Copenhagen region. Networks Spat Econ 2:327–346
Pelleg D, Moore AW, others (2000) X-means: extending k-means with efficient estimation of the number of clusters. In: Icml, pp 727–734
Rasouli S, Timmermans H (2016) Influence of social networks on latent choice of electric cars: a mixed logit specification using experimental design data. Networks Spat Econ 16:99–130
Robins G, Pattison P, Kalish Y, Lusher D (2007) An introduction to exponential random graph (p*) models for social networks. Soc Networks 29:173–191
Ross Morrow W, Long M, MacDonald EF et al (2014) Market-system design optimization with consider-then-choose models. J Mech Des 136:31003. https://doi.org/10.1115/1.4026094

Saberi M, Rashidi TH, Ghasri M, Ewe K (2018) A complex network methodology for travel demand model evaluation and validation. Networks Spat Econ 18:1051–1073
Sha Z, Huang Y, Fu S et al (2018) A Network-Based Approach to Modeling and Predicting Product Co-Consideration Relations. Complexity 2018
Sha Z, Saeger V, Wang M, Fu Y, Chen W (2017) Analyzing customer preference to product optional features in supporting product configuration. SAE Int J Mater Manuf 10:320–332
Shao W (2007) Consumer decision-making: an empirical exploration of multi-phased decision processes. Griffith University
Shocker AD, Ben-Akiva M, Boccara B, Nedungadi P (1991) Consideration set influences on consumer decision-making and choice: issues, models, and suggestions. Mark Lett 2:181–197
Stylidis K, Burnap A, Rossi M et al (2016) A preliminary study of trends in perceived quality design attributes in the automotive luxury market segment. In: DS 84: proceedings of the DESIGN 2016 14th international design conference, pp 2189–2196
Suryadi D, Kim HM (2019) A data-driven methodology to construct customer choice sets using online data and customer reviews. J Mech Des:1–19
Tanaka M, Ida T, Murakami K, Friedman L (2014) Consumers' willingness to pay for alternative fuel vehicles: a comparative discrete choice analysis between the US and Japan. Transp Res Part A Policy Pract 70:194–209
Tovares N, Cagan J, Boatwright P (2013) Capturing consumer preference through experiential conjoint analysis. ASME Pap No DETC2013–12549
Train K (1986) Qualitative choice analysis: theory, econometrics, and an application to automobile demand. MIT press
Train KE (2009) Discrete choice methods with simulation. Cambridge university press
Turrentine TS, Kurani KS (2007) Car buyers and fuel economy? Energy Policy 35:1213–1223
Wang M, Chen W, Fu Y, Yang Y (2015a) Analyzing and predicting heterogeneous customer preferences in China's auto market using choice modeling and network analysis. SAE Int J Mater Manuf 8:668–677
Wang M, Chen W, Huang Y, Contractor NS, Fu Y (2016a) Modeling customer preferences using multidimensional network analysis in engineering design. Des Sci 2
Wang M, Chen W, Huang Y et al (2015b) A Multidimensional Network Approach for Modeling Customer-Product Relations in Engineering Design. In: ASME 2015 International Design Engineering Technical Conferences & Computers and information in engineering conference, vol 2015, Boston
Wang M, Huang Y, Contractor N et al (2016b) A Network Approach for Understanding and Analyzing Product Co-Consideration Relations in Engineering Design. In: Proceedings of the DESIGN 2016 14th international design conference, Cavtat, Dubrovnik, Croatia, vol 2016
Wang M, Sha Z, Huang Y et al (2016c) Forecasting technological impacts on customers' co-consideration behaviors: a data-driven network analysis approach. In: ASME 2016 international design engineering technical conferences and computers and information in engineering conference, vol 2016, Charlotte, p V02AT03A040
Wang M, Sha Z, Huang Y et al (2018) Predicting Products' Co-Considerations and Market Competitions for Technology-Driven Product Design: A Network-Based Approach. Des Sci J 4
Wang P (2013) Exponential random graph model extensions: models for multiple networks and bipartite networks, Exponential random graph Model Soc networks Theory, methods, Appl, pp 115–129
Wang P, Pattison P, Robins G (2013a) Exponential random graph model specifications for bipartite networks' dependence hierarchy. Soc Networks 35:211–222
Wang P, Robins G, Pattison P, Lazega E (2013b) Exponential random graph models for multilevel networks. Soc Networks 35:96–115
Wasi N, Keane MP, others (2012) Estimation of Discrete Choice Models with Many Alternatives Using Random Subsets of the Full Choice Set: With an Application to Demand for Frozen Pizza
Wright P (1975) Consumer choice strategies: simplifying vs. optimizing. J Mark Res:60–67
Zhao M, Derudder B, Zhang P, Zhong P (2020) An expanded bipartite network projection algorithm for measuring cities' connections in service firm networks. Networks Spat Econ 20:479–498

Affiliations

Youyi Bi [1] · **Yunjian Qiu** [2] · **Zhenghui Sha** [3] · **Mingxian Wang** [4] · **Yan Fu** [4] · **Noshir Contractor** [5] · **Wei Chen** [6]

[1] University of Michigan-Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China

[2] Department of Aerospace & Mechanical Engineering, University of Southern California, 934 Downey Way, Los Angeles, CA 90089, USA

[3] Department of Mechanical Engineering, University of Arkansas, 863 W. Dickson St, Fayetteville, AR 72701, USA

[4] Global Data Insight & Analytics, Ford Motor Company, Dearborn, MI 48124, USA

[5] Department of Industrial Engineering & Management Sciences, Northwestern University, 2145 Sheridan Road, Tech C210, Evanston, IL 60208, USA

[6] Department of Mechanical Engineering, Northwestern University, 2145 Sheridan Road, Tech A216, Evanston, IL 60208, USA