2023 Conference on Systems Engineering Research

Network-Based Analysis of Heterogeneous Consideration-then-Choice Customer Preferences with Market Segmentations

Yaxin Cui[a], Yinshuang Xiao[b], Zhenghui Sha[b], Wei Chen[a] *

[a]*Department of Mechanical Engineering, Northwestern University, Evanston, IL, 60208, USA*
[b]*Department of Mechanical Engineering, University of Texas at Austin, Austin, TX, 78721, USA*

Abstract

Network-based analyses have been shown to be effective in understanding customer preferences by modeling the interactions and relationships between customers and products as a complex network system. Certain network representations, such as bipartite networks, can capture customers' two-stage (consideration-then-choice) decision-making processes by constructing either the "consideration" or "choice" links between the customer nodes and the product nodes. However, there is a dearth of research that examines network-based approaches to understanding complex heterogeneous customer preferences and the influence of product features in different market segments. This paper presents a market-segmented network modeling approach for understanding heterogeneous customer preferences in two-stage (consideration-then-choice) decision-making. Joint Correspondence Analysis is utilized to identify the correlations between product association networks and customer attributes, and then market segments are characterized by clustering customer attributes. We then construct bipartite customer-product networks and use the Exponential Random Graph Model to investigate factors that influence customer decision-making processes and how they vary among customer groups. A case study using real customer survey data for vacuum cleaners, a common household appliance with various product categories and a sizable market, serves to demonstrate the approach. The survey has been systematically designed and conducted on Cint platform to collect customer considerations and choices, product features, and customer attributes. Our findings reveal that customers are heterogeneous across different market segments which can be clustered based on their demographic attributes, usage contexts, and personal viewpoints. Within the identified market segments based on these aforementioned customer attributes, customer preferences toward product attributes show heterogeneity in different stages of choice-making. Particularly, it is observed that the majority of the product design attributes receive more attention in the consideration stage than in the choice stage. Our study advances the use of network-based models for analyzing customer preference heterogeneity across different market segments and in different stages of customers' decision-making.

Keywords: Heterogeneous customer preferences; Two-stage choice making; Complex networks; Network analysis; Market segments

* Corresponding author. Tel.: +1-847-491-7019.
E-mail address: weichen@northwestern.edu

1. Introduction

A quantitative understanding of customer preferences is vital to product design in many aspects, such as design attribute selection[1] and design optimization[2]. Network-based models have been increasingly used to quantitatively model customer preferences and behavior[3–6]. These models represent customers and products as nodes and their relations as edges in a network, allowing for the analysis of complex interactions and relationships between customers and products. Both exogenous attributes (e.g., product design features and customer attributes) and endogenous attributes (e.g., market structure effects) can be captured by various network models[7].

Meanwhile, there is a growing evidence in literature[8–11] on consumer research indicating that the complexity of customers' decision-making process consists of two different stages, i.e., consideration-then-choice, as shown in an illustrative example of vacuum cleaner purchase in Figure 1. The first stage involves a consideration decision process in which customers make initial selections of products to form a consideration set, while the second stage involves a compensatory process to derive the choice decision, where customers evaluate the tradeoffs among the product attributes. We have proposed a two-stage network-based modeling approach to study customers' consideration and choice behaviors[12], and the result suggested that the set of factors that influence customers' "consideration" of products in the first stage differ from the factors critical in the second stage of choice. However, this study was focused on one market segment (i.e., sedan) in the car market, where customers share similar needs and preferences, and customer heterogeneity was modeled by introducing customer attributes in addition to product attributes. In many consumer products markets, however, customer preferences are highly heterogeneous, meaning that besides the differences in social economic background and usage context, customers' preferences towards different product attributes can vary significantly among different customer groups. For instance, some customers may prefer products of high quality, while other customers may prefer more affordable products. When dealing with complex market with heterogeneities both within and across different market segments, a single network model that aggregates customer preferences towards each product attribute, either in a deterministic or statistical way[4], is not sufficient to understand and accurately model the complex heterogenous customer preferences. Therefore, further research is needed for modeling heterogeneous customer preferences in two-stage (consideration-then-choice) decision making and understanding the differences in product attributes influences for complex product markets with multiple market segmentations.

Fig. 1. Two-stage consider-then-choose decision-making in an example of customers purchasing vacuum cleaners. The two-stage choice model assumes that each customer considers a subset of products first and then makes the final decisions. Researchers have access to both consideration sets and the final choices data.

In market research, segmentation has been frequently employed for managing the complexity of modeling heterogeneous consumer preferences by classifying customers into homogeneous groups or segments with similar characteristics and needs[13,14]. It has been discovered that customer characteristics, including personal factors, psychological factors, and social factors, have a strong influence on their preferences[15]. Correspondingly, typical market segmentation techniques include demographic segmentation[16] (such as age, gender, and income), psychographic segmentation[16] (such as people's lifestyles and personal viewpoints), behavioral segmentation[17] (frequency of product usage), and need-based segmentation[18] (usage context). While market segmentation based customer preference modeling has been reported[19,20], existing works utilize conventional preference modeling techniques such as additive value functions[19] and discrete choice models[20], It remains a research topic of how to integrate market segmentations into network-based preference modeling, which is the main focus of this paper. Particularly, we intend to answer the following three **research questions**:

- RQ1: How can we identify market segmentations based on customer characteristics?
- RQ2: Can market-segmentation based network-based modeling lead to better results than using a single network model for the whole market?

- RQ3: What can we learn about the influence of product attributes in customers' two-stage consideration-then-choice decision-making process using market segmentation-based network modeling?

To answer these research questions, we propose a market segmentation-based network modeling approach to model heterogeneous customer preferences in two-stage decision making. In this approach, we first use Joint Correspondent Analysis (JCA) to visualize heterogeneous customer preferences and how they are associated with customer attributes (RQ1). Then cluster customers into different groups based on customers' attributes and construct bipartite networks consisting of customers and products to depict the customers' two-stage decision-making process of different customer groups. Lastly, the Exponential Random Graph Model (ERGM), a statistical network modeling approach, is utilized to investigate the important factors that influence customer consideration and choices in different market segments. A single network model without market segmentation serves as the baseline to be compared with our proposed model (RQ2 and RQ3).

Our approach is demonstrated using the data from vacuum cleaner customer survey, which was systematically designed to study multi-stage customer preference modeling in our recent work[21]. We choose the vacuum cleaner market as it is a common household appliance with heterogeneous categories and a large market size with diverse customers. The dataset contains 1,011 customer observations of 267 variables, including vacuum cleaner product attributes, customer purchase history (considered products and purchased products), and customer attributes (demographic attributes, usage context, and personal viewpoints).

2. Joint Correspondence Analysis for Customer Characteristics and Preferences

In this research, we first use joint correspondence analysis (JCA) as an exploration tool to identify the relationship between customer characteristics and their preferences. In this process, a product association network is first created using a unidimensional co-consideration network to find out the product communities that are more frequently co-considered by customers. Then JCA is used to relate product association communities to customer attributes to visualize the key customer attributes that drive what products they have considered. Here we use customers' considerations as an indicator of their preferences.

2.1 Technical background of JCA

Joint correspondence analysis (JCA) is a statistical technique that is used to analyze and visualize the relationship between multiple categorical variables. JCA has been widely used in marketing research, for studying the relationship between consumers and products[22]. For example, JCA was used to identify groups of consumers who have similar purchasing patterns[23], or to determine products that are most likely to be purchased together[22]. It has also been used in the social sciences to study the relationship between demographic variables and attitudes or behaviors[24]. In network-based customer preference modeling[25], it was introduced as a multivariate approach for graphical data representation, which offers a visual understanding of the connections between product consideration sets and the relations with customer attributes. Originated from correspondence analysis (CA), a method for analyzing two-way contingency tables based on the singular value decomposition of a matrix of correspondence weights, JCA extends it to allow for the analysis of multiple contingency tables, or joint distributions, which can provide a more comprehensive view of the relationships between variables. In essence, JCA is a dimension reduction method to visualize the data matrix in a subspace of low dimensionality.

2.2 Product community detection

Distinct customer preference types, which can be denoted by the different products they have considered, need to be identified to investigate the relationship between customer preferences and their characteristics. To achieve this, we first build a product association network that reveals product communities based on their co-consideration relationships. In the survey data, each customer reported the product they have considered, which can reveal product co-consideration relations. For example, "Dyson Upright Vacuum Cleaner, Ball Multi Floor 2" and "Dirt Devil Razor Pet Bagless Upright Vacuum" have been co-considered by customers in our survey data, so there will be a co-consideration link between them. In network analysis, a community refers to a group of nodes with denser connections internally and sparser connections between the groups. By detecting the communities in a product co-consideration network, we reveal the groups of products that are strongly co-considered with others, treat them as different market segments, and then further identify the customer preferences for each market segment.

Table 1. Product community detection and characteristics (mean value and (standard deviation)) of each community

	Communities	Characteristics	Dominant type	Suction power (rating)	Price (dollar)	Weight (lb.)	Representative model
	Community 1	High-tech and expensive	Robotic	2.62 (1.25)	296.12 (205.22)	8.73 (5.09)	
	Community 2	Traditional and affordable	Upright	2.90 (1.20)	179.37 (166.19)	11.68 (6.36)	
	Community 3	Strong suction power	Upright	2.99 (1.34)	259.85 (218.60)	10.20 (6.78)	
	Community 4	Innovative and portable	Stick	2.75 (1.22)	228.47 (201.72)	8.77 (5.76)	

Network community detection uses algorithms developed from graphical properties. In this research, the Spinglass algorithm[26] has been used. With elbow rules, we detect the optimum size of communities is 4. In Table 1, we provide the community detection result and summarize the key characteristic with the mean value and standard deviation of each community. The product features of each community are used to summarize the characteristics of each community. Community 1 contains the most expensive vacuum cleaners, and the most dominant type is robotic vacuum cleaners, which may be summarized as “High-tech and expensive”. Community 2 contains the vacuum cleaners with lowest price and the most dominant type is upright vacuum cleaners, which can be named as “Traditional and affordable”. “Strong suction power” is used to represent Community 3's highest suction power, while "Innovative and portable" describes Community 4's majority of lighter-weight stick vacuums. The detected communities serve as different types of customer preferences as they are more frequently co-considered by customers.

2.3 Visualization of JCA results of customer attributes and product communities

Based on detected product communities, we use JCA to explore the correlation between customer attributes and their preferences for vacuum cleaner models from different communities. We have included varieties of customer attributes in this analysis, such as demographic attributes (income, education level), usage context (house type, number of rooms), and personal viewpoints (attitudes towards innovation, environmentally friendly, price sensitive, quality). Here in JCA, an indicator matrix was constructed that each row is a customer observation, and the column variables are their considered vacuum cleaner models and customer attributes in categories.

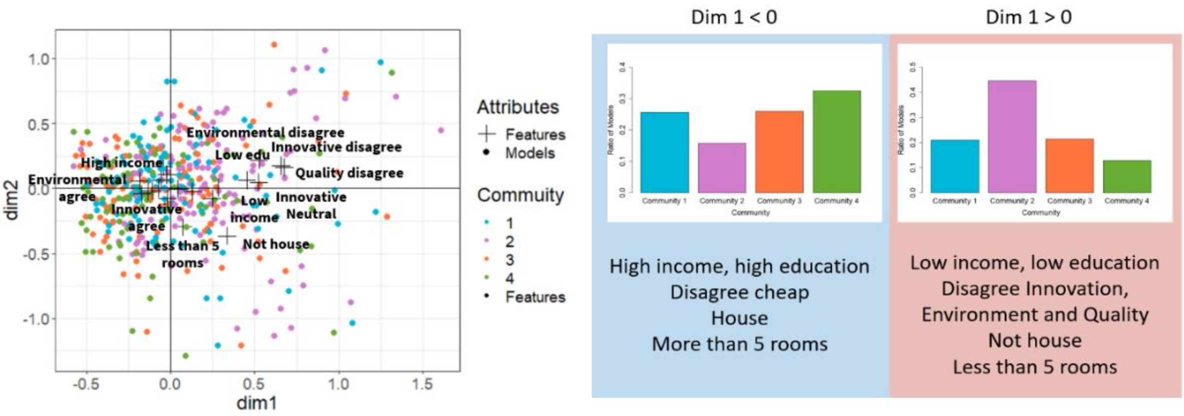

Fig. 2. Joint correspondence analysis based on vacuum cleaner communities and customer attributes: original algorithm results (left) and illustration of the vacuum cleaner type distribution on dim 1 < 0 and dim 1 >0 (right). In the region of dim 1 < 0, most products are from community 1, 3, and 4, and in the region of dim > 0, most products are from community 2.

In Fig. 2, with the reduced dimensionality representation using JCA, the distance between two points can be interpreted as relative similarities in the variables examined. The dots in the plot denote vacuum cleaner models that have been considered by customers with four different colors corresponding to four different communities, and the crosses represent customer attributes. There are three main aspects of interpretations of the plot. First, if two products are considered by customers with similar profiles, they are closer to each other in the plot. We observe that vacuum cleaner models from community 1 (high-tech and expensive) and 4 (innovative and portable) are closer to each other in JCA visualization, which means they are frequently considered by customers with similar profiles. Products from community 2 in magenta (traditional and affordable) are frequently considered by similar customers as most of them are close to each other, and 62% of them are distributed in the region of dim1 > 0. Second, two customer attributes are closer if they often appear together for specific customers. It is notable that in the region of dim1 > 0, the customer features with lower education level and income are closer to customers who don't live in a house (non-house) and with less than 5 rooms. Meanwhile, those attributes are much closer to customers who hold neutral or negative attitudes to innovation, environment friendly and quality and prefer cheaper products compare to those who hold positive attributes. Third, a product and a demographic attribute are placed close to each other if customers consider the product often possesses the attributes. We observe that customers with a more price-sensitive attribute are closer to the products that are more traditional and affordable, and customers with more enthusiasm for innovation, environment friendly and quality have more preference for high-tech and lower suction power models in communities 1 and 4.

In summary, from the JCA plot, we observe that different product communities are strongly correlated with different customer attributes. There are at least two major categories of customers ("price-sensitive" and "innovation-passionate"), divided by dim1 > 0 versus dim1 < 0, respectively, whose preferences are very different as they choose vacuum cleaners associated with different product communities. This observation serves as the justification for us to determine market segments based on customer attributes in this research.

3. Bipartite Consideration and Choice Networks

3.1. Customer segmentations

Based on the visualization of customer attributes and their associations with product communities in the JCA plots, we conclude that customer attributes (demographic attributes, usage context and personal viewpoints) can be used as the base for market segmentation. In this paper, we use k-mode methods to cluster customers[27] into two separate clusters, using the same customer attributes considered in JCA presented in Section 2.

Some representative customer features in each cluster are plotted in Figure 3. We observe the profile of customers from cluster 1 ("innovation-passionate") is 75.7% with high income, 40.3% with high education, 69.5% has more than five rooms, 87.1%, 77.9%, and 86.1% respectively agree that innovation, environmentally friendly features and quality are very important in their decision-making process. On the other hand, the profile of customers from cluster 2 ("price-sensitive") is 33.4% with high income, 35.0% with high education, 62.9% has more than five rooms, 75.3%, 55.2%, and 68.6% respectively agree that innovation, environmentally friendly features and quality are very important. We expect customers from different clusters (market segments) would have different preferences towards product attributes in their vacuum cleaner purchase.

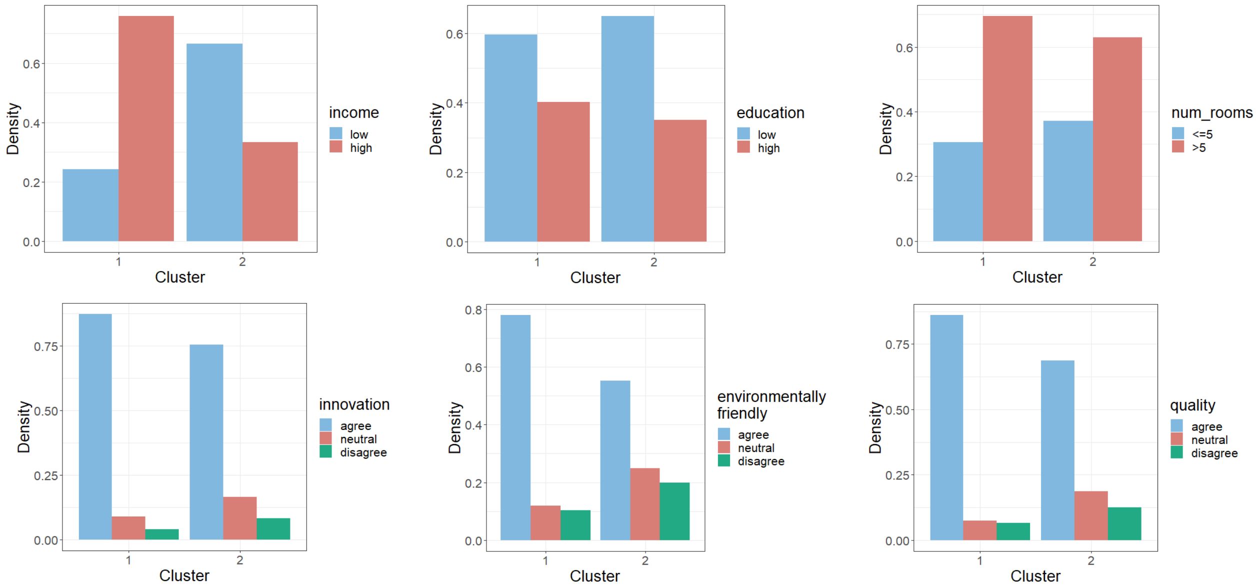

Fig. 3. Customer attributes distribution in each cluster. Customers from cluster 1 are the innovative-passionate type, and customers from cluster 2 are the price-sensitive type.

3.2. Construction of consideration and choice networks

As shown in our previous research, the customer-product relationship can be represented by a bipartite network[12,28], where customers and products are modeled as two types of nodes, and the considerations and choices of the customers are modeled as different types of links. Therefore, customer decision making in consideration-then-choice can be viewed as modeling the likelihood of forming consideration or choice links between nodes. In stage 1, consideration links in the bipartite network represent customers' considerations among all products, and in stage 2, choice links in the bipartite network indicate the final purchase decision among all products being considered, conditional on the consideration set from stage 1.

In this paper, based on the identified market segments using customer clusters, we separate 1,011 vacuum cleaner customers into two different groups and construct separate bipartite networks to analyze customer preferences towards product attributes. The customer-product bipartite networks are plotted in Figure 4, where black dots are customers and colored dots are products belonging to different communities. As the consideration sets include multiple products, but choice is made only for one project, the network for stage 1(consideration) is much denser than that for stage 2 (choice), and the links in the choice network are conditional on the consideration stage. The node sizes for vacuum cleaners represent their popularity (how frequently considered or chosen) in the given network. For customers in cluster 2 (price-sensitive), the products from community 2 (traditional and affordable) are more prevalent in both the consideration stage and the choice stage.

For customers (innovation-passionate) in cluster 1, there are 380 customers and they have considered 408 vacuum cleaner models. The density of the consideration network is 0.0031 and that of the choice network is 0.0012. For customers (price-sensitive) in cluster 2, there are 491 customers and they have considered 397 vacuum cleaner models. The density of the consideration network is 0.0027, and that of the choice network is 0.0012. These data shows that the two networks decomposed based on market segmentation have comparable size and density.

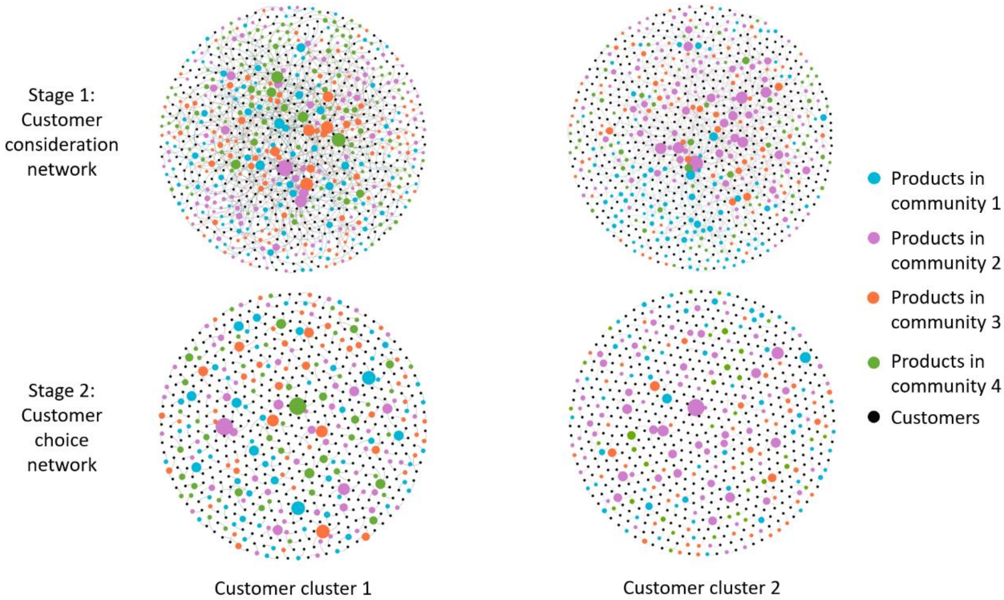

Fig. 4. Bipartite consideration and choice networks in different customer clusters (market segments)

4. Network-based models for analyzing customer preferences towards product attributes

4.1. Exponential Random Graph Models

Recent advances in Exponential Random Graph Models (ERGMs) provide a unified and flexible statistical inference framework for multidimensional network analysis[29]. ERGM takes a network as one entity and allows researchers to model the interdependence (or network structure effects) among nodes[30]. For example, if nodes tend to connect with those nodes that already have many links in a network, they exhibit a star-type structure. The key idea behind ERGM is that it considers an observed network, y, as one specific instance from a set of possible random networks, $\boldsymbol{Y}$, following the distribution in Equation (1)

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

where $\boldsymbol{\theta}$ is a vector of model parameters ($\boldsymbol{\theta}'$ is the transpose of $\boldsymbol{\theta}$), $g(\boldsymbol{y})$ is a vector of the network statistics, and $\kappa(\boldsymbol{\theta})$ is a normalizing factor to ensure Equation (1) is a proper probability distribution. Equation (1) suggests that the probability of observing any network is proportional to the exponent of a weighted combination of network characteristics: one statistic $g(\boldsymbol{y})$ is more likely to occur if the corresponding $\boldsymbol{\theta}$ is positive. It is worth noting that in ERGMs, the network itself is a random variable and the probability is evaluated on the entire network instead of a link. ERGM uses Markov Chain Monte Carlo (MCMC) simulation to estimate the parameter values $\boldsymbol{\theta}$ that maximize the likelihood of observed network structures at the aggregate level.

Bipartite ERGM is a type of ERGM specifically for modeling two-level network structures[31,32], and following the same model structure as defined in Equation (1) except that $g(\boldsymbol{y})$ only captures network statistics within the bipartite affiliation network. Unlike one-mode networks which consist of a single layer of nodes (e.g., social influence network and co-consideration network), the bipartite network is comprised of links connecting two layers (two types of nodes). In this work, there are two types of links in two separate bipartite networks: consideration links and choice links.

4.2. Modeling setting for two-stage modeling

To model the binding relations of two-stage decision-making, we need to set constraints to make sure that the choice stage is conditional on the consideration stage that customers choose products within their choice sets. There are two specific settings to mimic the real choice scenarios in the ERGM setting. The first is a constraint to control that only product nodes connected in the consideration stage can be linked to each customer in the choice stage. The second is a constraint to set the limit that each customer node only links to one product node so that each customer can only make one choice from the consideration set.

4.3. Results of two-stage modeling results

Table 2. ERGM results of two-stage modeling results for different customer clusters (market segmentations)

	Customer cluster 1: innovation passionate type		Customer cluster 2: Price sensitive	
Model terms	Stage 1 (consideration)	Stage 2 (choice)	Stage 1 (consideration)	Stage 2 (choice)
Edges	-6.5754***	N/A	-6.2397***	N/A
Market distribution	0.5021*	-0.3428*	-0.0133	0.0260
Upright vacuum	0.1843*	0.3780.	0.5654***	0.7179**
Robotic vacuum	0.1248*	-0.0989	0.2559	-0.2373
Price	0.0008***	0.0002	0.0002	0.0007
Filter: HEPA	0.2564**	-0.1657	0.2348**	-0.0718
Capacity	0.0241***	0.0216*	0.0228**	0.0210
Bagless	0.6827***	0.1770	0.4157**	0.4890
Suction power	0.0994***	-0.1183	0.0544.	-0.1277.
Model fit: AIC	11725	-3.118	13323	-2.798
Model fit: BIC	11815	62.33	13415	78.65

Note. .p < .1; *p < .05; **p < .01; ***p < 0.001

Table 2 summarizes the estimation of product attributes and network structural effects in the two-stage decision-making process for two distinct customer clusters. There is no estimation for the “Edges” term in the choice stage as there is a fixed number of edges in the choice network where each customer only chooses one product. All the numerical product attributes are normalized to the range of [0,1] before we cluster the customers.

The network structural effects include *edges* (which can be regarded as a constant in the model) and market distribution. The *market distribution* is measured with an endogenous structural variable in ERGM – geometrically weighted degree distribution (GWDegree)[33], which describes the distribution of the degree of vacuum cleaner models. A positive coefficient of *market distribution* indicates an even degree distribution (i.e., most vacuum cleaners have similar numbers of sales), while a negative coefficient indicates a skewed degree distribution (i.e., a few vacuum cleaners have much bigger sales than others). We notice that for the consideration stage for customer cluster 1, the market distribution shows a positive and significant effect, which implies that most vacuum cleaners have similar chances of being considered by customers. However, for the choice stage for customer cluster 1, the market is more skewed distributed, and a few vacuum cleaner models are more dominant in the market.

The estimation of nodal attributes denotes the importance of vacuum cleaner product attributes in the customer consideration and choice stages. It is observed that in **the consideration stage**, for both customer clusters, the influence of most product features (*HEPA filter*, *capacity*, *bagless*, and *suction power*) are significantly significant. Based on the estimated results for customer clusters 1 and 2, vacuum cleaners with *HEPA filters*, larger *capacity*, *bagless*, and larger *suction power* are more preferred in the consideration stage, and those features shows more significance for customers in cluster 1 than those in cluster 2. Customers from cluster 2 (price-sensitive type) have a stronger preference towards *upright vacuum cleaners*, while those from cluster 1 (innovation-passionate type) prefer *robotic vacuum cleaner* more than customers in cluster 2. Also, we notice the difference in *price* influence on customers in different segments: the customers from cluster 1 (innovation-passionate) prefer more expensive vacuum cleaners which implies better quality and more advanced technologies, while the price does not have an important effect for customers from cluster 2 (price-sensitive).

We also observe that in **the choice stage**, product features are less influential in customer decision-making compared to the consideration stage, given that only *capacity* in customer cluster 1 and *suction power* in customer cluster 2 have marginally significant effects. This could be because of the existing model's inadequacy in accounting for other aspects (such as customer ratings and online recommendations) that might be essential during the choice stage, and most product features have already been taken into account during the consideration stage.

By comparing the key factors in the customer decision-making process in different stages and for different customer groups, we can see the differences of preferences between the two different customer clusters. Customers in "innovation-passionate" cluster focus more on the product features and are willing to pay more for better products, whereas customers in "price-sensitive" cluster have a stronger preference for *upright vacuum cleaners* with more traditional designs. Additionally, product features have greater effects on the consideration stage and less impact during the choice stage.

We also run the baseline single network model without market segmentation, and the estimated parameters for ERGM terms in each stage are recorded in Table 3. The single bipartite network contains data from 871 customers involving 528 products. We notice that the obtained coefficient of market distribution (-1.1262) in the single network at the consideration stage is negative, meaning that the market distribution is more skewed, and some products are more frequently considered than the rest of the products in the market. On the contrary, with market-segment based network modeling, the products are more evenly considered by customers in customer cluster 1. For preference towards vacuum cleaner type, the single model shows that both upright and robotic vacuum cleaners are preferred compared to other types of vacuum cleaners but can't differentiate the specific product attributes that are influential for each type. The market-segment based models overcome this difficulty and we can observe the differences in preferences for upright and robotic vacuum cleaners for different customer clusters in Table 2. In stage 2, the results for the single network model show that *upright vacuum cleaners* are strongly preferred in the choice stage (which is true because upright vacuum cleaners have the largest market share), but we found that if we separate customers into clusters, this is only true to the customer cluster 2. Therefore, the results could be misleading when we try to understand customers' preferences without market segmentation. Also, we notice that *capacity* has a positive effect and suction power has a negative effect in the choice stage. However, in the model with market segmentation, *capacity* only exerts a positive effect on customer cluster 1 and *suction power*'s negative effect is only significant for customer cluster 2. While we have reported the AIC and BIC values for the model fit measurement, we cannot compare it directly with the network models with market segmentation, as those metrics are only comparable for the same model structure and input data.

Table 3. ERGM results of two-stage modeling results for a single network model (without market segmentation)

Model terms	Stage 1 (consideration)	Stage 2 (choice)
Edges	-6.2055***	N/A
Market distribution	-1.1262***	-0.3042.
Upright vacuum	0.3439***	0.5073**
Robotic vacuum	0.1859**	-0.1162
Price	0.0004***	0.0004
Filter: HEPA	0.1879**	-0.1097
Capacity	0.0173***	0.0246*
Bagless	0.4180***	0.3076
Suction power	0.0552**	-0.1145*
Model fit: AIC	26090	-15.29

Model fit: BIC	26189	73.20

5. Conclusion

In this research, we propose a market segmentation-based network modeling approach for identifying heterogeneous customer preferences in the two-stage consideration-then-choice decision-making process. More specifically, we first used the Joint Correspondence Analysis to visualize the relationship between product association communities and customer attributes (demographic, usage context, and personal viewpoints) in a latent space with a reduced dimensionality. The process allows us to identify market segments characterized by clustering of customer attributes. For each identified market segment, we construct bipartite customer-product networks which denote the customer-product relations in the customers' consideration and choice stages respectively, while the choice stage network is conditional on the consideration stage to mimic customers' decision-making process. Finally, by adapting the Exponential Random Graph Model we investigate the different factors that influence customer decision-making processes and how they are different for distinct customer groups. Using the real customer survey data collected through Cint platform for vacuum cleaner, the results indicate that product attributes play more important roles in the consideration stage compared to the choice stage, and the same product attributes could have different effects on different market segments (innovation-passionate customers versus price-sensitive customers in our case study). While only capacity (for innovation-passionate customers) and suction power (for price-sensitive customers) shows marginally significant results in the choice stage. The network-based model for different market segmentations can interpret customer preference more practically compared to using a single network. To the authors' knowledge, this paper is the first study to investigate customer segmentation in network-based customer preference modeling for understanding complex heterogeneous customer preferences. It provides a data-driven approach for identifying market segmentations instead of making arbitrary assumptions. It lays the groundwork for future research into more comprehensive network-based methods in exploring customer preferences in different product markets.

While our research has shed light on exploring heterogeneous customer preference by partitioning customers into different groups in the form of network analysis using the co-consideration network for detecting product associations, there are several limitations and future directions of the current work. First, the current models focus on product-related features and market distribution effects. We could explore more interaction effects between customers and products and investigate whether there is a strong influence of interaction effects, especially in the choice stage. Second, we currently only evaluate our model based on the model convergence (models are converged using Markov chain Monte Carlo simulation) and goodness of fit (AIC and BIC) based on the network model, but there is no direct comparison between each the model performance. There are other potential mechanisms of validating the model by making predictions of customer choices which would be explored in the future. Also, the current market segmentation is based on clustering using a small set of customer attributes, which can be expanded, and the different clustering approaches can be employed.

Acknowledgements

The authors acknowledge collaborators, Noshir Contractor, Neelam Modi, Jonathan Haris Januar, Johan Koskinen, Michael Cardone, Gracia Cosenza, and Emma Jackson, for their assistance in survey design and data processing and the inputs and feedbacks during research meetings. We also greatly acknowledge the funding support from NSF CMMI #2005661 and #2203080.

References

1. Hoyle, C. J. & Chen, W. Product attribute function deployment (PAFD) for decision-based conceptual design. *IEEE Trans. Eng. Manag.* **56**, 271–284 (2009).
2. Wassenaar, H. J. & Chen, W. An approach to decision-based design with discrete choice analysis for demand modeling. *J. Mech. Des.* **125**, 490–497 (2003).
3. Wang, M., Chen, W., Huang, Y., Contractor, N. S. & Fu, Y. Modeling customer preferences using multidimensional network analysis in engineering design. *Des. Sci.* **2**, (2016).

4. Wang, M., Chen, W., Fu, Y. & Yang, Y. 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 (2015).
5. Sha, Z. *et al.* Modeling product co-consideration relations: A comparative study of two network models. in *Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 6: Design Information and Knowledge, Vancouver, Canada, 21-25.08. 2017* (2017).
6. Cui, Y. *et al.* A Weighted Statistical Network Modeling Approach to Product Competition Analysis. *Complexity* **2022**, (2022).
7. Sha, Z. *et al.* Comparing Utility-based and Network-based Approaches in Modeling Customer Preferences for Engineering Design. *Proc. Des. Soc. Int. Conf. Eng. Des.* **1**, 3831–3840 (2019).
8. Shocker, A. D., Ben-Akiva, M., Boccara, B. & Nedungadi, P. Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Mark. Lett.* **2**, 181–197 (1991).
9. Shao, W. *Consumer Decision-Making: An Empirical Exploration of Multi-Phased Decision Processes*. (Griffith University, 2007).
10. Hauser, J. R., Ding, M. & Gaskin, S. P. Non-compensatory (and compensatory) models of consideration-set decisions. in *2009 Sawtooth Software Conference Proceedings, Sequin WA* (2009).
11. Gaskin, S., Evgeniou, T., Bailiff, D. & Hauser, J. Two-Stage Models: Identifying Non-Compensatory Heuristics for the Consideration Set then Adaptive Polyhedral Methods Within the Consideration Set. (2007).
12. Fu, J. S. *et al.* Two-Stage Modeling of Customer Choice Preferences in Engineering Design Using Bipartite Network Analysis. in *IDETC-CIE2017* (2017). doi:10.1115/DETC2017-68099.
13. Beane, T. P. & Ennis, D. M. Market Segmentation: A Review. *Eur. J. Mark.* **21**, 20–42 (1987).
14. Goyat, S. The basis of market segmentation: a critical review of literature. *Eur. J. Bus. Manag.* **3**, 45–54 (2011).
15. Consumer behavior in marketing - patterns, types, segmentation - Omniconvert Blog. *Omniconvert Ecommerce Growth Blog* https://www.omniconvert.com/blog/consumer-behavior-in-marketing-patterns-types-segmentation/ (2019).
16. Lin, C. Segmenting customer brand preference: demographic or psychographic. *J. Prod. Brand Manag.* **11**, 249–268 (2002).
17. Susilo, W. H. An Impact of Behavioral Segmentation to Increase Consumer Loyalty: Empirical Study in Higher Education of Postgraduate Institutions at Jakarta. *Procedia - Soc. Behav. Sci.* **229**, 183–195 (2016).
18. Peltier, J. W. & Schribrowsky, J. A. The use of need-based segmentation for developing segment-specific direct marketing strategies. *J. Direct Mark.* **11**, 53–62 (1997).
19. Liu, J., Liao, X., Huang, W. & Liao, X. Market segmentation: A multiple criteria approach combining preference analysis and segmentation decision. *Omega* **83**, 1–13 (2019).
20. Modeling Preference and Structural Heterogeneity in Consumer Choice. https://pubsonline.informs.org/doi/epdf/10.1287/mksc.15.2.152 doi:10.1287/mksc.15.2.152.
21. Xiao, Y. *et al.* Information Retrieval and Survey Design for Two-Stage Customer Preference Modeling. *Proc. Des. Soc.* **2**, 811–820 (2022).
22. Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research - Donna L. Hoffman, George R. Franke, 1986. https://journals.sagepub.com/doi/abs/10.1177/002224378602300302.
23. Beldona, S., Morrison, Alastair. M. & O'Leary, J. Online shopping motivations and pleasure travel products: a correspondence analysis. *Tour. Manag.* **26**, 561–570 (2005).
24. de Nooy, W. Fields and networks: correspondence analysis and social network analysis in the framework of field theory. *Poetics* **31**, 305–327 (2003).
25. Wang, M., Huang, Y., Contractor, N., Fu, Y. & Chen, W. *A Network Approach for Understanding and Analyzing Product Co-consideration Relations in Engineering Design*. (2016).
26. Reichardt, J. & Bornholdt, S. Statistical Mechanics of Community Detection. *Phys. Rev. E* **74**, 016110 (2006).
27. Chaturvedi, A., Green, P. E. & Caroll, J. D. K-modes Clustering. *J. Classif.* **18**, 35–55 (2001).
28. Bi, Y. *et al.* Modeling Spatiotemporal Heterogeneity of Customer Preferences in Engineering Design. in *ASME 2018 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference* (2018).
29. Handcock, M. S. *et al.* Package 'ergm'. (2015).
30. Butts, C. T. *et al.* Introduction to Exponential-family Random Graph (ERG or p*) modeling with ergm. *Eur. Univ. Inst. Florence URL Httpcran R-Proj. Orgwebpackagesergmvignettesergm Pdf* (2014).
31. Wang, P. Exponential Random Graph Model Extensions: Models for Multiple Networks and Bipartite Networks. in *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications* (eds. Lusher, D., Robins, G. & Koskinen, J.) 115–129 (Cambridge University Press, 2012). doi:10.1017/CBO9780511894701.012.
32. Wang, P. Exponential random graph model extensions: Models for multiple networks and bipartite networks. *Exponential Random Graph Models Soc. Netw. Theory Methods Appl.* 115–129 (2013).
33. Hunter, D. R. Curved Exponential Family Models for Social Networks. *Soc. Netw.* **29**, 216–230 (2007).