

# DATA-DRIVEN DYNAMIC NETWORK MODELING FOR ANALYZING THE EVOLUTION OF PRODUCT COMPETITIONS

**Jian Xie**

Industrial Engineering Laboratory, Beijing Institute of Technology  
5 South Zhonguancun St., Haidian Dist., Beijing China 100081  
[xiejian@bit.edu.cn](mailto:xiejian@bit.edu.cn)

**Youyi Bi**

University of Michigan-Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong  
University, 800 Dong Chuan Road, Shanghai, China, 200240  
[youyi.bi@sjtu.edu.cn](mailto:youyi.bi@sjtu.edu.cn)

**Zhenghui Sha**

System Integration & Design Informatics Laboratory, University of Arkansas  
863 W. Dickson St., Fayetteville, AR, USA 72701  
[zsha@uark.edu](mailto:zsha@uark.edu)

**Mingxian Wang**

Global Data Insight & Analytics, Ford Motor Company  
Dearborn, MI, USA  
[MingxianWang2016@u.northwestern.edu](mailto:MingxianWang2016@u.northwestern.edu)

**Yan Fu**

Global Data Insight & Analytics, Ford Motor Company  
Dearborn, MI, USA  
[yfu4@ford.com](mailto:yfu4@ford.com)

**Noshir contractor**

Science of Networks in Communities, Northwestern University  
Evanston, IL, USA  
[nosh@northwestern.edu](mailto:nosh@northwestern.edu)

**Lin Gong**

Industrial Engineering Laboratory, Beijing Institute of Technology  
5 South Zhonguancun St., Haidian Dist., Beijing China 100081  
[gonglin@bit.edu.cn](mailto:gonglin@bit.edu.cn)

**Wei Chen<sup>1</sup>**

Integrated Design Automation Laboratory, Northwestern University  
2145 Sheridan Road, Tech A216, Evanston, IL 60208-3109  
[weichen@northwestern.edu](mailto:weichen@northwestern.edu)

---

<sup>1</sup> Corresponding author, Wilson-Cook Professor in Engineering Design, Northwestern University  
Email: [weichen@northwestern.edu](mailto:weichen@northwestern.edu)

## **ABSTRACT**

Understanding the impact of engineering design on product competitions is imperative for product designers to better address customer needs and develop more competitive products. In this paper, we propose a dynamic network-based approach to modeling and analyzing the evolution of product competitions using multi-year buyer survey data. The product co-consideration network, formed based on the likelihood of two products being co-considered from survey data, is treated as a proxy of products' competition relations in a market. The Separate Temporal Exponential Random Graph Model (STERGM) is employed as the dynamic network modeling technique to model the evolution of network as two separate processes: link formation and link dissolution. We use China's automotive market as a case study to illustrate the implementation of the proposed approach and the benefits of dynamic network models compared to the static network modeling approach based on Exponential Random Graph Model (ERGM). The results show that since STERGM takes preexisting competition relations into account, it provides a pathway to gain insights into why a product may maintain or lose its competitiveness over time. These driving factors include both product attributes (e.g., fuel consumption) as well as current market structures (e.g., the centralization effect). With the proposed dynamic network-based approach, the insights gained from this paper can help designers better interpret the temporal changes of product competition relations to support product design decisions.

**Keywords:** engineering design, product competition, dynamic network analysis, ERGM, STERGM

## 1. INTRODUCTION

In a dynamic market environment, product competition changes over time [1], e.g., due to release of new products, withdrawal of outdated products, technological progress, and social changes. For example, the market share of U.S. brands in the Chinese auto market fell from 12.2% in 2017 to 10.7% in 2018, and according to [2–4] a possible reason for this decline is due to the delayed refresh of the U.S. lineups, which implies the negative effect of untimely design improvement on product competition. The change of customer preferences is another potential cause for the evolution of product competition. Fuel-efficient cars have been more desirable since the energy crisis of the 1970s as customers become more sensitive to rising gas prices [5]. Since early 2000s, higher fuel efficiency has contributed to the increasing competitiveness of hybrid vehicles [6]. Therefore, a thorough understanding of the dynamic changes in market competition is of great significance in many engineering design scenarios, such as product feature competition (e.g., whether to upgrade design features of an existing car and by how much) and new product positioning (e.g., whether to develop a new car model to fill a specific market niche).

Existing research on product competition has primarily focused on investigating the strategic factors that influence the competitiveness of products. For example, Talay et al. [7] employed the parametric hazard model to investigate the factors influencing the survival of auto companies in the U.S. automotive market. They found that rather than the quality of its innovations, it is a firm's ability to keep up with its competitors that determines its chances of survival in the marketplace. This again demonstrates the significance of investigating the competition relations of products. Roberts [8] applied the autoregressive profit model to examine the relationship between product innovation, competition and profitability in the U.S. pharmaceutical industry. They found the positive relationship between a firm's ability to avoid competition (i.e., to sustain the competitive

position of each innovation over time) and the persistence of its above-normal profit outcomes. Economic growth model [9] and game theoretic models [10,11] have been adopted for understanding the dynamics of product competition. However, these methods are not able to directly model the influence of engineering attributes and existing competition relations on future product competitions. Thus, they may not provide sufficient insights into the impact of design improvement on maintaining competition relations.

To understand the impact of product design change on competition relations, recent studies have explored the use of network analysis to model product competitions [12,13]. Network-based approaches model products as nodes and the co-consideration relations (competition) between products as links. Here, the product co-consideration network is treated as a proxy of a competition network in a market. By taking the dependencies among links into consideration, network-based approaches are effective in modeling complex and interdependent relations [14,15]. In previous research, a unidimensional network-based approach based on Exponential Random Graph Model (ERGM) [16] has been developed to study product competitions in the form of product co-considerations [13], and used to assess the impact of technological changes (e.g., turbo engine) on product competitions and customers' co-consideration behaviors [17]. Later, the unidimensional network-based approach was extended to a multidimensional network structure where customer social relations were included to study their influence on heterogeneous customer preferences [15]. Despite the strength of these earlier developed network models [12,13,15,17–19], they are static in nature that ignore the dynamic change of product competitions over time. The models are limited in explaining why a product may maintain or lose its competitive advantage over time. To overcome the limitations of existing methods, we propose a dynamic network modeling approach to study the evolution of product competition relations. Specifically, we utilize the Separable

Temporal ERGM (STERGM) [20,21] for dynamic network modeling based on the following reasons:

- First, existing degree-based generative dynamic network models (e.g., small world network [22], scale-free network [23], and the dynamic stochastic block model [24]) are limited in modeling the influence of nodal and edge characteristics (e.g., customer demographics and relation strength) on the change of network structures [25]. In contrast, inferential techniques such as STERGM [20] examines both growing and shrinking dynamics, i.e., how the explanatory factors influence the formation of new links and the dissolution of old links. Compared to the Stochastic Actor-Oriented Model (SAOM) [26], another widely used longitudinal network model which treats nodes as actors who make active decisions, STERGM is more advantageous for our purpose because it is a link-oriented model that can include nodes which are not actors, e.g., products in our work [27].
- Second, in the social network literature, STERGM has been utilized for modeling various dynamic social relations, such as the evolutions of social networks of politicians [28] and international trade networks [29]. Researchers have found that both exogenous factors (e.g., economic characteristics of countries) and preexisting network structures (e.g., reciprocity for bilateral trade relations and triadic closure effects for trilateral trade relations) influence international trade networks over time [29]. Thus, STERGM allows us to investigate how preexisting competition relations influence product competitions in addition to the impact of product features.
- Third, STERGM is an extension to the static ERGM. Therefore, the results from STERGM and ERGM can be compared and related to further advance our understanding of product competition relations.

However, STERGM has never been employed for studying dynamic product competitions. Our objective in this paper is to extend STERGM from social network analysis to studying the impact of engineering design and existing market competitions on the evolution of product competition relations. A data-driven approach is developed to create the STERGM using multi-year buyer survey data. The network links are formed based on the likelihood of two products being co-considered using the choice set information gathered from buyers. Here co-consideration is a situation that a customer concurrently considers multiple products in cross-shopping activities [30]. We choose co-consideration to represent competitiveness in this research because a product will not be purchased by customers if not considered first.

In existing literature, STERGM only models the same set of nodes over time in dynamic networks [28,29]. However, for product competitions, the sets of products (or nodes) in different years vary because new products appear in the market and existing products exit the market from time to time. To overcome this difficulty, we leverage the concept of “structural zeros” [31] to address the varying product consideration set from year to year in dynamic network modeling. Results from STERGM are used to explain what factors drive two products into competition and why a competitive relationship is preserved or dissolved over time. The insights gained can help designers develop more competitive product thus better address customer needs. They may also support companies’ strategic decision making, such as branding, product positioning, and marketing.

The paper is organized as follows. Section 2 provides the technical background of network analysis, the static network modeling technique (ERGM), the dynamic network modeling technique (STERGM), and the “structural zeros” concept. In Section 3, a general approach for modeling dynamic product competition relations based on STERGM is introduced and illustrated

using the data associated with China's SUV market. The process of preparing the dataset, identifying modeling attributes, and handling varying vehicle nodes in dynamic network modeling are explained. The results of STERGM are presented and compared with the results of the static network modeling approach (ERGM). To assess the results, model fit evaluation is performed at both the link level and the network level. Section 4 demonstrates how the proposed approach can predict future product competitions compared to ERGM using PR (Precision-Recall) curve and support decision-making in vehicle design. Section 5 concludes with reflections on the implications of this approach for engineering design and future research opportunities.

## **2. TECHNICAL BACKGROUND OF DYNAMIC NETWORK ANALYSIS AND MODELING**

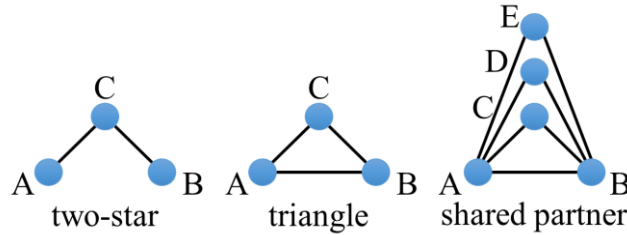
Network analysis has been recognized as an essential method for analyzing and modeling complex systems in a wide variety of fields such as biology, computer science, social science, and engineering [15,32–34]. With network analysis, the structure of a system is visualized and simplified as a graph, where nodes represent entities in the system and edges/links represent relationships between entities. In this section, two statistical network models for network analysis, i.e., Exponential Random Graph Model (ERGM) and Separable Temporal Exponential Random Graph Model (STERGM), and the “structural zeros” concept are introduced.

### **2.1. Exponential Random Graph Model (ERGM)**

ERGM is a flexible statistical inference framework, which assumes an observed network  $y$  as an instance of random networks  $Y$  given by the distribution in Eqn. (1),

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

where  $\theta$  is a vector of corresponding model parameters, and  $\kappa(\theta, \mathbf{y})$  is a normalizing constant to guarantee the equation is a proper probability distribution. In the context of this work, network  $\mathbf{Y}$  captures product competition relations, which are identified based on product co-considerations from the customer survey (see more details in Section 3.1).  $\mathbf{g}(\mathbf{y})$  is a vector of network statistics of interest including attributes of nodes, attributes of links as well as network structure attributes [15]. Fig. 1 provides three exemplary network structures that can be modeled by ERGM. Eqn. (1) suggests that the probability of observing a specific network structure is proportional to the exponent of a weighted combination of network statistics. The estimated ERGM parameters  $\theta$  indicate the importance of the network statistics to the formation of links in a network. For example, a positive  $\theta$  of the shared partner structure in an ERGM for the vehicle competition network implies that if two vehicles compete with some other vehicles, they are more likely to compete with each other.



**FIGURE 1: Exemplary network structures in a vehicle competition network.** The two-star structure indicates whether a car will have two distant competitors. The triangle structure indicates whether three cars compete with each other (three-way competition). The shared partner structure indicates whether two cars sharing the same set of competitors will be co-considered.

The estimated ERGM parameters  $\theta$  can also be used to calculate the log-odds of the formation of certain links, i.e., how likely a link exists between two nodes given their nodal attributes and the structure of the remaining network as shown in Eqn. (2):

$$\text{Logit } \Pr(y_{ij} = 1 | \mathbf{y}_{-ij}) = \log \frac{\Pr(y_{ij}=1 | \mathbf{y}_{-ij})}{\Pr(y_{ij}=0 | \mathbf{y}_{-ij})} = \theta^T \cdot (\mathbf{g}(\mathbf{y} | y_{ij} = 1) - \mathbf{g}(\mathbf{y} | y_{ij} = 0)) = \theta^T \cdot \delta_{ij}(\mathbf{y}), \quad (2)$$



where  $y_{ij}$  is the link between node  $i$  and  $j$ ,  $\mathbf{y}_{-ij}$  is the network excluding the link between node  $i$  and  $j$ , and  $\delta_{ij}(\mathbf{y})$  is the difference of the network statistics between the network where the link between node  $i$  and  $j$  exists (i.e.,  $\mathbf{g}(\mathbf{y}|y_{ij} = 1)$ ) and the network where the link between node  $i$  and  $j$  does not exist (i.e.,  $\mathbf{g}(\mathbf{y}|y_{ij} = 0)$ ). If we get positive log-odds for nodes  $i$  and  $j$ , this indicates that having a link between them is more likely than not having the link and *vice versa*.

## 2.2. Separable Temporal Exponential Random Graph Model (STERGM)

As an extension of ERGM, STERGM is established to model dynamic networks. Shown in Fig. 2, STERGM treats the evolution from the network at time  $t$  ( $\mathbf{Y}^t$ ) to the network at time  $t + 1$  ( $\mathbf{Y}^{t+1}$ ) as two separate processes: 1) **link formation** in which new links are created following  $\Pr(\mathbf{Y}^+ = \mathbf{y}^+ | \mathbf{Y}^t; \boldsymbol{\theta}^+)$ , and 2) **link dissolution** in which old links following  $\Pr(\mathbf{Y}^- = \mathbf{y}^- | \mathbf{Y}^t; \boldsymbol{\theta}^-)$ . Here  $\boldsymbol{\theta}^+$  and  $\boldsymbol{\theta}^-$  denote the parameters of the formation model ( $\mathbf{Y}^+$ ) and the dissolution model ( $\mathbf{Y}^-$ ), respectively. The network at time  $t + 1$  is constructed by applying the changes in  $\mathbf{Y}^+$  and  $\mathbf{Y}^-$  to  $\mathbf{Y}^t$  following Eqn. (3):

$$\mathbf{Y}^{t+1} = \mathbf{Y}^- \cup (\mathbf{Y}^+ - \mathbf{Y}^t). \quad (3)$$

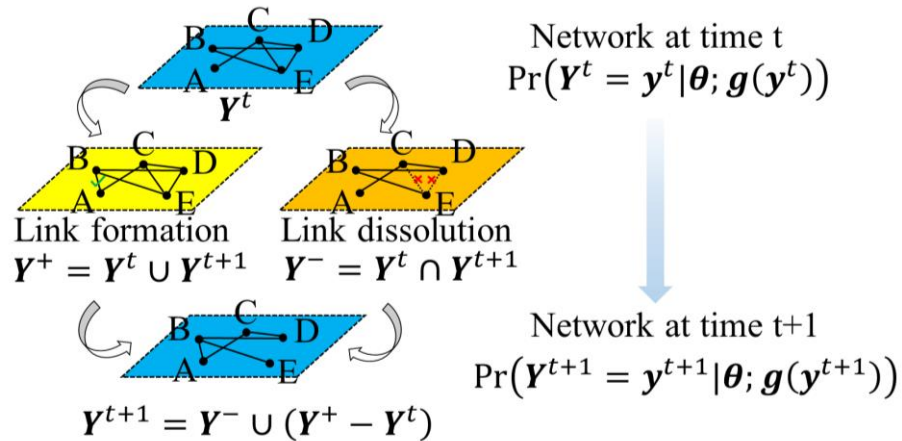


FIGURE 2: Evolution dynamics of product competition network

For each discrete time step, the process of formation and dissolution are independent conditional on the network at time  $t$ . This means,

$$Pr(Y^{t+1} = \mathbf{y}^{t+1} | Y^t = \mathbf{y}^t; \boldsymbol{\theta}) = Pr(Y^+ = \mathbf{y}^+ | Y^t = \mathbf{y}^t; \boldsymbol{\theta}^+) \cdot Pr(Y^- = \mathbf{y}^- | Y^t = \mathbf{y}^t; \boldsymbol{\theta}^-), \quad (4)$$

Relating STERGM with ERGM, the probability distribution of formation model and dissolution model are expressed as

$$Pr(Y^+ = \mathbf{y}^+ | Y^t = \mathbf{y}^t; \boldsymbol{\theta}^+) = \frac{\exp(\boldsymbol{\theta}^{+T} \cdot \mathbf{g}(\mathbf{y}^{t+1}, \mathbf{y}^t))}{\kappa(\boldsymbol{\theta}^+, \mathbf{y}^t)}, \quad (5) \text{ and}$$

$$Pr(Y^- = \mathbf{y}^- | Y^t = \mathbf{y}^t; \boldsymbol{\theta}^-) = \frac{\exp(\boldsymbol{\theta}^{-T} \cdot \mathbf{g}(\mathbf{y}^{t+1}, \mathbf{y}^t))}{\kappa(\boldsymbol{\theta}^-, \mathbf{y}^t)}. \quad (6)$$

The normalizing denominator  $\kappa(\boldsymbol{\theta}^+, \mathbf{y}^t)$  and  $\kappa(\boldsymbol{\theta}^-, \mathbf{y}^t)$  are the sum of network statistics of all possible formation and dissolution networks, respectively. Here these formation and dissolution networks only include possible variations to  $\mathbf{Y}^t$  (i.e., the additions and subtractions). In contrast, the normalizing denominator of ERGM includes all networks from an empty network to a complete network (i.e., a network in which all nodes are linked with each other).

Similar to ERGM, STERGM can include both exogenous (e.g., nodal attributes) and endogenous (network structures) variables in network modeling. This enables the prediction of products' future competition relations considering the market's present competition structure and the influence of design changes. In addition, not only can STERGM identify the design features contributing to the formation of competition between two products, but it can also identify the features influencing the dissolution of competitions. For example, when assessing the influence of certain SUV-relevant attributes on vehicles' competitiveness given the current market structure, if the estimated effect of third-row seat feature is positive and statistically significant in the formation model, it implies that improving the design by including third-row seats would make the SUV more likely to be considered against its competitors. In the dissolution model, if the estimated parameter of fuel

consumption is negative and statistically significant, it means that better fuel economy in the SUV would make its competing relationships more likely to not dissolve (that is, persist).

### 2.3. Structural zeros

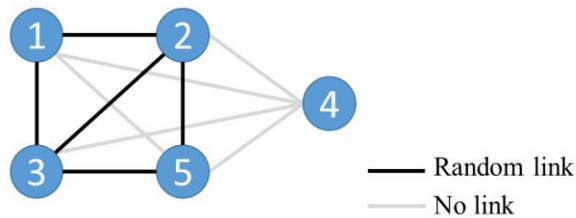
STERGM is typically used to model the evolution of dynamic networks with the same set of nodes at different time points. This is not suitable for modeling competitive dynamics among a set of products that vary from one year to another (i.e., inconsistent set of nodes). To address this challenge, we utilize the concept of *structural zeros*. Given a unidimensional network consisting of  $N$  nodes, this network can be presented by a  $N \times N$  matrix with entries of binary indicator 0 or 1. Structural zeros are a way to predefine zero inputs in the matrix to restrict those relations among nodes in the network that are logically incapable of having links [35]. For instance, a car introduced at time  $t+1$  cannot logically have a competitive relationship with another car that existed at time  $t$ . Therefore, the distribution of dynamic network at time  $t + 1$  can be written as

$$Pr(\mathbf{Y}^{t+1} = \mathbf{y}^{t+1} | \mathbf{Y}^t = \mathbf{y}^t; \mathbf{R}^t = \mathbf{r}^t; \boldsymbol{\theta}), \quad (7)$$

where  $\mathbf{R}^t \in \mathbb{R}^{N \times N}$  is binary matrix corresponding to  $\mathbf{Y}^t$  and  $\mathbf{r}^t$  is a realization of  $\mathbf{R}^t$ , such as

$$\mathbf{r} = \begin{bmatrix} \mathbf{0} & 1 & 1 & \mathbf{0} & \mathbf{0} \\ 1 & \mathbf{0} & 1 & \mathbf{0} & 1 \\ 1 & 1 & \mathbf{0} & \mathbf{0} & 1 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & 1 & \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad (8)$$

where blocks denoted as  $\mathbf{0}$  are structural zero blocks which enforce  $\mathbf{y}_{15}$ ,  $\mathbf{y}_{51}$ ,  $\mathbf{y}_{x4}$ ,  $\mathbf{y}_{4x}$  and  $\mathbf{y}_{xx}$  ( $x = 1, 2, \dots, 5$ ) to be 0 in  $\mathbf{Y}^{t+1}$ , while the other entries denoted as 1 indicate corresponding



**FIGURE 3: A random network example incorporating structural zeros shown in Eqn. (8)**

relations in  $\mathbf{Y}^{t+1}$  have no restrictions (i.e., they are free to change to 0 or remain as 1). Fig. 3 shows an exemplary network following this matrix.

Thus, the dynamic network model based on structural zeros can be represented as

$$Pr(\mathbf{Y}^{t+1} = \mathbf{y}^{t+1} | \mathbf{Y}^t = \mathbf{y}^t; R^t = r^t; \boldsymbol{\theta}) = \frac{\exp(\boldsymbol{\theta}^{+T} \cdot \mathbf{g}(\mathbf{y}^{t+1}, \mathbf{y}^t))}{\kappa(\boldsymbol{\theta}^+, \mathbf{y}^t, r^t)} \cdot \frac{\exp(\boldsymbol{\theta}^{-T} \cdot \mathbf{g}(\mathbf{y}^{t+1}, \mathbf{y}^t))}{\kappa(\boldsymbol{\theta}^-, \mathbf{y}^t, r^t)}, \quad (9)$$

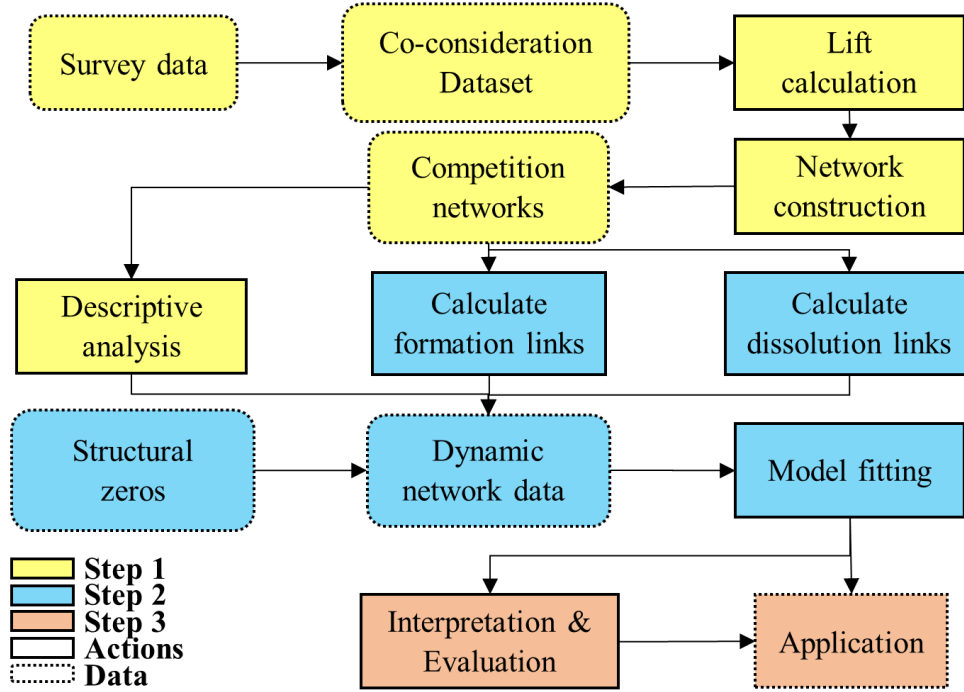
where possible networks of  $\mathbf{Y}^+$  and  $\mathbf{Y}^-$  in  $\kappa(\boldsymbol{\theta}^+, \mathbf{y}^t, r^t)$  and  $\kappa(\boldsymbol{\theta}^-, \mathbf{y}^t, r^t)$  have been reduced because of  $r^t$ .

The ability to restrict the value of certain cells in a matrix using structural zeros has been applied in different areas, such as parametric models [36], control systems [37] and social networks [38,39]. For example, in social network area, Robins et.al. [38] used structural zeros to represent the relationship of paired persons in a romantic relation network that are not closely associated romantically. In customer preference modeling, Fu et al. [19] and Sha et al. [40] used structural zeros to constrain customers from purchasing products outside their consideration sets in a bipartite network-based approach for modeling customer preferences over two stages (consideration and purchase). Snijders et al. [26] used structural zeros to deal with the missing data in friendship networks. In this paper, we employ this concept to handle the issue of nodes that are not present across all time periods.

### 3. UTILIZING STERGM TO MODEL EVOLUTION OF PRODUCT COMPETITIONS

#### 3.1. Overview of the dynamic network analysis approach

Fig. 4 illustrates the procedure of utilizing STERGM for analyzing the evolution of product competitions in market. The detailed description of each of the three steps is provided as follows.



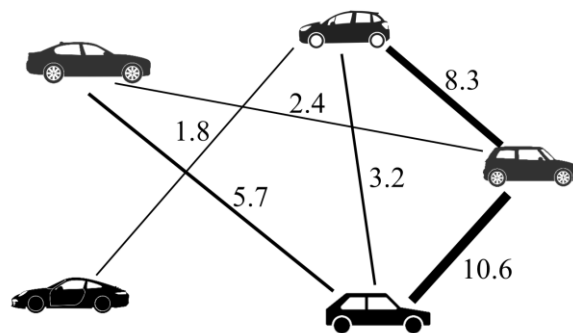
**FIGURE 4: Overview of the proposed approach**

Step 1- *Network construction and descriptive analysis*. Product competition relationship is identified in this work based on the survey data of co-considerations, i.e., the products in the same consideration set before a customer makes the final choice. In the competition network, whether the competition relationship exists is determined by the *lift* criterion [18]. As one of the standard metrics of association rules [41], *lift* can be used to measure the strength of the association between two products based on how often they appear in the same consideration set. The calculation of *lift* and the criterion for link formation are illustrated in Eqn. (10),

$$\text{Lift}(i, j) = \frac{Pr(i, j)}{Pr(i) \cdot Pr(j)}, \quad y_{ij} = \begin{cases} 1, & \text{if } \text{lift}(i, j) > \text{threshold} \\ 0, & \text{otherwise} \end{cases}, \quad (10)$$

where  $Pr(i, j)$  is the probability of products  $i$  and  $j$  being co-considered by customers,  $Pr(i)$  is the probability of individual product  $i$  being co-considered with other products, and  $y_{ij}$  is the

competition link between product  $i$  and product  $j$ . The *lift* value measures the likelihood of competition between two products given their respective frequencies of considerations and indicates the dependence of the two products being considered. If the *lift* between product  $i$  and product  $j$  is greater than a threshold, then the competition link between these two product nodes exists. The threshold is determined based on whether the formed network sufficiently addresses the purpose of study, robustness analysis as well as the convergent requirements of the model. In this paper, we set the threshold equals to 1 to capture two products that compete more likely than expected in random [13]. The robustness analysis using threshold at 1, 3, 5, and 7 were conducted and no significant changes in network characteristics and the trend of model results are observed. Fig. 5 provides an illustrative unidimensional undirected vehicle competition network, in which the numbers represent the *lift* values. A larger *lift* value between two vehicles indicates that they have higher likelihood being co-considered by customers, i.e., there is a stronger competition relationship between them. Descriptive analysis and visualization of the obtained network can provide an intuitive understanding of the nodal attributes and network structures.



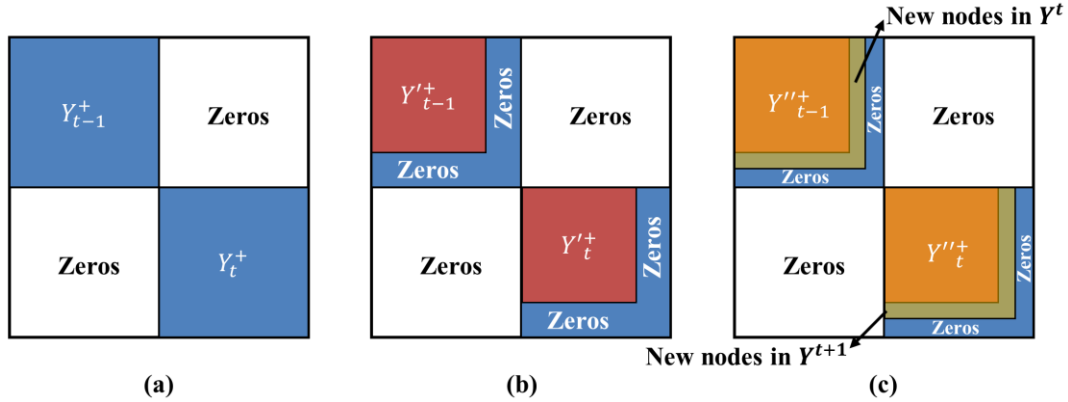
**FIGURE 5. An illustrative vehicle competition network.** Each link represents a competition relationship between two cars. The numbers are the *lift* values. The thickness of a link is proportional to the value of *lift*.

**Step 2- Dynamic network modeling.** After generating the product competition networks in different time periods, chosen as one year per period, we can construct the formation and dissolution

networks between each pair of two temporally consecutive networks (see Fig. 2), and use STERGM to investigate the evolution of network structures. The results of descriptive analysis in Step 1 can help the selection of explanatory factors in network statistics  $g(\mathbf{y}^{t+1}, \mathbf{y}^t)$ . For example, an obvious trend or significant change of attributes in different years and in different vehicles may imply the significance of influence. These attributes may be considered as explanatory factors in the next step. In addition, the selection of explanatory factors also depends on a modeler's interest. To deal with the varying nodes, structural zeros were utilized in STERGM to generate the dynamic network model (see Eqn. 8). Fig. 6 shows how this method works for link formation process by using three networks, represented by adjacency matrices, in time  $t-1$ ,  $t$  and  $t+1$  (i.e.,  $\mathbf{Y}_{t-1}, \mathbf{Y}_t, \mathbf{Y}_{t+1}$ ) as an example. The formation network between  $\mathbf{Y}_{t-1}$  and  $\mathbf{Y}_t$  is noted as  $\mathbf{Y}_{t-1}^+$ , and the formation network between  $\mathbf{Y}_t$  and  $\mathbf{Y}_{t+1}$  is noted as  $\mathbf{Y}_t^+$ . The structural zeros method contains the following steps:

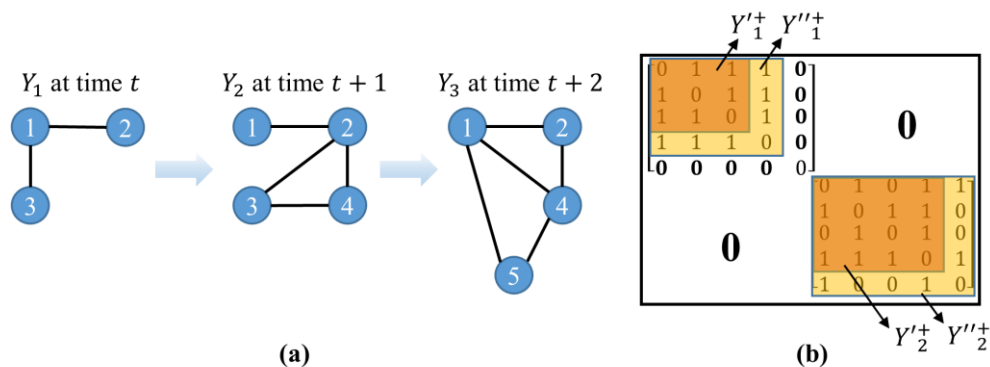
- a) Combine all possible formation networks ( $\mathbf{Y}_{t-1}^+$ ,  $\mathbf{Y}_t^+$  in this example) in one matrix. If we need to model four-year networks, there will be three formation networks, i.e., three blue regions in Fig. 6 (a). Since STERGM only supports consistent sets of nodes in principle, here the matrices of  $\mathbf{Y}_{t-1}^+$  and  $\mathbf{Y}_t^+$  include the union set of nodes from  $\mathbf{Y}_{t-1}, \mathbf{Y}_t, \mathbf{Y}_{t+1}$ , which means the sizes of these two sub-matrices are the same as shown in the blue regions in Fig. 6 (a). Apparently, the whole matrix is symmetric with zero diagonal cells (self-competitions are not considered).
- b) Fill zeros to specific blue regions of Fig. 6 (a) to ensure those new nodes appearing in the other years have no links. For example, the red region  $\mathbf{Y}'_{t-1}$  in Fig. 6 (b) only includes the vehicles appearing in  $\mathbf{Y}_{t-1}$ , and  $\mathbf{Y}'_t$  only includes the vehicles appearing in  $\mathbf{Y}_t$ .

- c) To ensure that the model captures the effect for both preserved and created nodes, the new nodes in the second year of a two-year formation network which are not considered in the red zone in Fig. 6 (b) will be included in the yellow zone in Fig. 6 (c). For example, the difference between the yellow region  $Y''^+_{t-1}$  in Fig. 6 (c) and the red region  $Y'^+_{t-1}$  in Fig. 6 (b) is the new nodes appearing in  $Y_t$ .



**FIGURE 6: Structural zeros used in STERGM for dynamic network with varying sets of nodes.** (a) Combining all possible formation networks in one adjacency matrix. (b) Utilizing structural zeros to restrict the relations between nodes in current year and new nodes in other years. (c) Releasing the restriction of new nodes in the second year to capture the effect for created nodes.

Upon completion, the matrix in Fig. 6 (c) will replace the original matrix in Fig. 6 (a) in fitting the formation model of STERGM. Fig. 7 gives an example of how to use this method in a dynamic



**FIGURE 7: An illustrative example for structural zeros method in formation model.** (a) A dynamic network example containing five nodes 1-5 and three networks  $Y_1$ ,  $Y_2$  and  $Y_3$  in three time periods  $t$ ,  $t+1$  and  $t+2$ . (b) Corresponding adjacency matrix of formation networks based on (a), in which red regions represent Fig. 6 (b) and yellow regions represent



network containing three time periods where the bolded zeros represent the structural zeros in the formation model. Here we focus on the formation process for illustrative purpose, and the dissolution process follows the same procedures except that the dissolution networks are used.

*Step 3- Interpretation, evaluation and application.* STERGM can be estimated by methods commonly used in ERGM such as maximum likelihood and generalized moments [29]. In this paper, Conditional Maximum Likelihood Estimation (CMLE) was utilized to model the transition between two networks and estimate the coefficients in STERGM. The estimation results should be evaluated based on both model fitting and model interpretability. In addition, the coefficients of product attributes are interpreted from a dynamic perspective. The model is evaluated and verified in both link-level and network-level to make sure the results are credible. Insights gleaned from how product competitions change over time can be used to support product upgrades and new product development strategies in engineering design. In this paper, we apply our approach to predicting future product competitions and conversely examining the influence of engineering attributes on the competitiveness of a product.

### **3.2. Dataset for a Case Study on China's Crossover SUV Market**

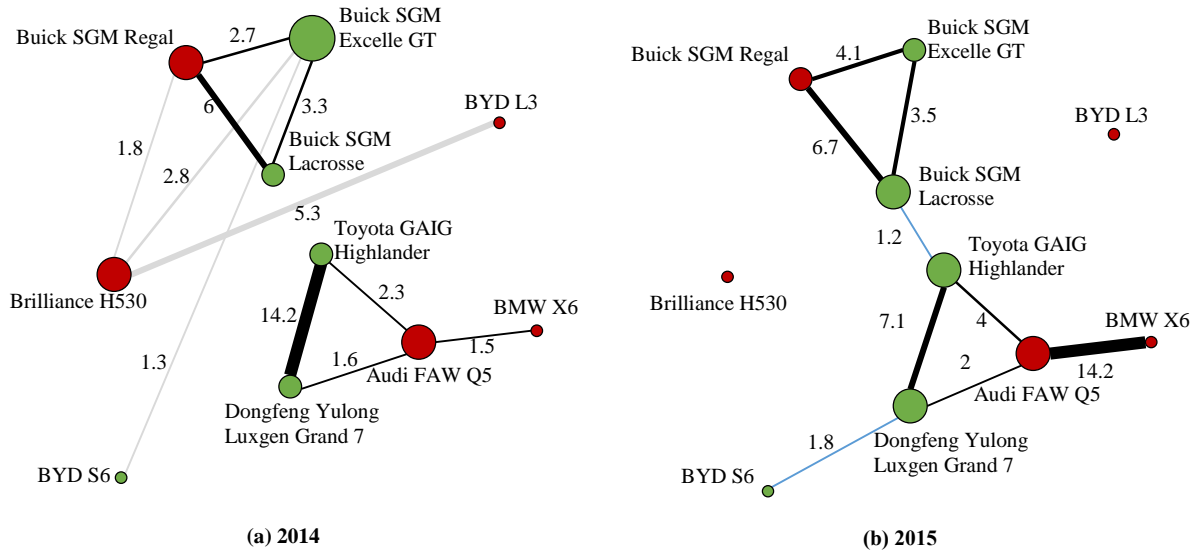
To demonstrate the proposed approach, we carried out a case study using the data from a recognized new car buyer survey from 2013 to 2015 in China's auto market [19]. The dataset in each year consists of 50,000-70,000 new car buyers' responses about approximately 400 unique car models. Respondents were asked to list up to three vehicles including final choice they considered when purchasing a car. By using each individual buyer's consideration set, the *lift* value between any pair of car models can be calculated following Eqn. (10) based on which the co-consideration network is constructed. Customer demographics, such as income, city of residence and education, and car attributes, such as price, power, and fuel consumption, are reported in the

survey and verified by the company. In each year, the survey was collected every four months. However, we use yearly data in this paper to avoid the seasonal effects on customers' choice behaviors. We followed previous research in organizing the market survey dataset. More details of data collection, cleaning and preprocessing can be found in [13,15,18,42]. In this case study, we focus on market competitions among crossover SUVs. A crossover SUV, such as an Audi Q7, BMW X6, or Ford Edge, is a vehicle with the body and space of an SUV, but made using a platform shared with sedan. Crossover SUVs inherit the advantages of spacious interiors from vans, the off-road performance of SUVs and the light-weight and fuel economy of sedans. Our interest is to demonstrate the factors that influence two crossover SUVs entering and ceasing a competitive relationship.

### 3.3. Descriptive analysis

We utilize the customers' consideration reported in the survey data to calculate the *lift* value (see Eqn. 10) for all competitions between car models. We constructed three competition networks corresponding to years 2013, 2014 and 2015. There are 115, 139 and 119 car models from 2013 to 2015 considered respectively in our study (including crossover SUVs and other conventional or SUV vehicles co-considered with crossover SUVs) and 600 pairs of competitions in each year. Fig. 8 illustratively shows a vignette of a partial competition networks in 2014 and 2015 with 10 vehicle models. The size of a node is proportional to its degree and the nodes with green color represent car models that were selected for log-odds calculation in Section 3.5 (this is to verify the STERGM modeling results at link levels). The blue line indicates newly formed competitive links in 2015 while the grey lines indicate the dissolved ones. The *lift* value marked on each link shows the strength of the competition and indicates the dependence degree of two competing products. For example, in Fig. 8, Audi FAW Q5 (Q5) competes with three cars - Toyota GAIG Highlander

(Highlander), Dongfeng Yulong Luxgen Grand 7 (Luxgen Grand 7), and BMW X6 (X6). For a customer considering the Q5, in 2014 the Highlander is the most likely co-considered car whereas in 2015 the closest competitor changes to X6.



**FIGURE 8: An example of partial vehicle competition networks evolving from 2014 (a) to 2015 (b).** Black lines indicate preserved competitions, grey lines indicate dissolved competitions, and blue lines represent new competitions.

Table 1 provides the mean value and standard deviation (in brackets) of vehicle attributes (except ratings for seat position and legroom) from 2013 to 2015 for all pairs of competing vehicles versus all possible pairs of competing vehicles. The major difference is that the car models in the set of “all possible pairs of competing vehicles” do not necessarily hold a direct competition relation (e.g., car A and car B are not competitors, but they both compete with car C). Based on our previous research in choice modeling and product competitions [12,15,43], three categories of vehicle attributes are considered in modeling: 1) regular vehicle attributes such as the price, power and fuel consumption of vehicles; 2) comparative vehicle attribute, reflecting their difference. For example, the price difference between two competed vehicles allows us to investigate whether vehicles with similar or different attributes are more likely to compete with each other; and 3)

SUV-relevant attributes, such as seat position. In total, nine vehicle attributes are considered including price, power, fuel consumption (FC), turbo, make origin, all-wheel drive (AWD), seat position, legroom, and third row. Among them, price and power are preprocessed using log2 transformation to handle their non-normal distributions. Turbo, AWD, and third row are binary variables describing whether such an attribute is available in a car or not. Seat position and legroom are based on customer satisfaction Likert-scale ratings from 1 (dissatisfy strongly) to 4 (satisfy strongly). Median values were used in Table 1 for seat position and legroom since it is more appropriate than mean value for a Likert ordinal scale (then the standard deviation is not needed) [44,45]. Table 1 shows that the mean values of price, power, fuel consumption, vehicle attributes difference, and third row are lower on average in competing vehicles than that in all possible pairs of competing vehicles. Some vehicle attributes changed from 2013 to 2015; for example, the mean value of power for all possible pairs of competing vehicles increased from 7.25 to 7.35 and fuel consumption decreased from 10.13 to 9.96.

**Table 1. Descriptive analysis of competing vehicle pairs versus all possible pairs of competing vehicles**

	Pairs of competing vehicles			All possible pairs of competing vehicles		
	2013	2014	2015	2013	2014	2015
<b>Regular vehicle attributes</b>						
Price(yuan)	16.95	17.37	17.04	17.43	17.55	17.52
(log2)	(0.94)	(1.22)	(0.99)	(1.22)	(1.37)	(1.32)
Power(BHP)	7.03 (0.42)	7.20 (0.52)	7.18 (0.41)	7.25 (0.56)	7.28 (0.58)	7.35 (0.55)
(log2)						
FC (L per 100 km)	9.24 (1.79)	9.82 (2.07)	9.36 (1.73)	10.13 (2.3)	10.07(2.17)	9.96 (2.09)
Turbo	0.14 (0.28)	0.28 (0.39)	0.37 (0.37)	0.19 (0.33)	0.28 (0.39)	0.39 (0.40)
Origin (US)	172 (14%)	239 (16%)	138 (15%)	23 (20%)	21 (15%)	18 (15%)
Origin (EU)	188 (15%)	366 (25%)	111 (12%)	20 (17%)	36 (26%)	27 (23%)
Origin (JP)	183 (15%)	167 (11%)	107 (12%)	19 (17%)	17 (12%)	18 (15%)
Origin (KR)	275 (22%)	128 (9%)	59 (7%)	17 (15%)	14 (10%)	10 (8%)
Origin (CN)	438 (35%)	566 (39%)	485 (54%)	36 (31%)	51 (37%)	46 (39%)
<b>Vehicle attribute difference</b>						

Price difference	0.47 (0.39)	0.49 (0.43)	0.44 (0.36)	1.40 (1.01)	1.57 (1.13)	1.52 (1.09)
Power difference	0.25 (0.21)	0.25 (0.22)	0.27 (0.21)	0.64 (0.47)	0.67 (0.49)	0.63 (0.46)
FC differences	1.08 (0.94)	1.08 (0.94)	1.04 (0.86)	2.61 (1.95)	2.48 (1.80)	2.42 (1.71)
<b>SUV-relevant attributes</b>						
AWD*	0.14 (0.29)	0.26 (0.36)	0.22 (0.30)	0.23 (0.36)	0.26 (0.37)	0.24 (0.36)
Seat position	3	3	3	3	3	3
Legroom	3	3	3	3	3	3
Third row	0.10 (0.29)	0.14 (0.35)	0.12 (0.33)	0.16 (0.36)	0.16 (0.37)	0.20 (0.40)

\* AWD: all-wheel drive

### 3.4. Results of Dynamic Network Modeling

The R package “tergm” is used to fit the STERGM models [46]. As shown in Table 2, the explanatory variables in our dynamic network models correspond to two types of variables: endogenous variables (i.e., network structures) and exogenous variables consisting of the main effects and homophily effects of vehicle attributes. The main effect measures the impact of the existence or value of a vehicle attribute on the probability of a competition link, whereas the homophily effect measures the impact of the similarity (or difference) of the attributes of two vehicles on the probability of a competition link [42]. In this study, we consider three network structure statistics [47]: Edges, which estimates the likelihood that two cars will compete with each other randomly. It can be treated as a “base rate” similar to the intercept term in a regression. Geometrically Weighted Edgewise Shared Partner (GWESP), which is referred to as the shared partner structure in Fig. 1 and measures the closure effect of the network; and Geometrically Weighted Degree (GWD), which measures the centralization effect of the network (i.e., the evenness of degree distribution). In this study, the decay parameter for GWESP and GWD was fixed to 0.5 for simplifying the computation. In a vehicle competition network, a positive coefficient of GWESP means it is very likely for two cars to compete with each other if they share

the same set of competitors. A positive coefficient of GWD means most cars have similar numbers of competitors. A negative coefficient of GWD means some cars are much more likely to have a much larger number of competitors than the rest. As such it reflects a “centralization effect” in the competition network. The Akaike information criterion (AIC) [48] was used as the estimator to measure the relative quality of the network model in this paper. Lower AIC value indicates better quality of a model.

Three STERGM models (A, B, C) are created with different model specifications. In Model A, we only consider main effect of regular vehicle attributes and homophily effects, and results are used as the baseline for comparison. Model B includes main effect of regular vehicle attributes, homophily effects, and network effects. In Model C, we consider all endogenous and exogenous variables including main effect of SUV-relevant attributes. The reduced AIC value from Model A to Model B in Table 2 indicates model improvement due to the introduction of network structure effects. This means that network structure effects play an important role in forming new competitions and preserving existing competitions. Further, the improvement in model quality from Model B to Model C indicates the inclusion of SUV-relevant attributes is essential in explaining the formation of competitions. However, introducing SUV-relevant attributes did not incrementally explain the dissolution of competitive links as evidenced in the AIC of Model C compared to Model B. Overall, Model C is considered as the best fitted model.

**Table 2. Results of the STERGM fitting for dynamic competition networks from 2013 to 2015**

Model	Formation			Dissolution		
	A <sup>1</sup>	B <sup>2</sup>	C <sup>3</sup>	A	B	C
<b>Network effects</b>						
Edges	5.83***	-1.55	-1.91	9.35	7.11	-2.08
Closure		0.88***	0.85***		0.41**	0.40**

Centralization		-0.77**	-0.66*		0.26	0.50
<b>Main attributes effects</b>						
Price	-0.16	0.01	-0.12	-0.28	-0.18	-0.36
Power	-0.18	-0.11	-0.10	-0.24	-0.35	-0.33
Fuel consump.	0.08*	0.03	0.05	0.22	0.21	0.16
Turbo	0.24**	0.11	0.11	0.06	-0.07	-0.11
Make origin(US)	-0.03	-0.15	-0.10	0.29	0.33	0.65*
Make origin(EU)	0.10	-0.03	0.03	0.58	0.44	0.68
Make origin(JP)	-0.07	-0.18	-0.09	0.09	-0.09	0.06
Make origin(KR)	-0.31**	-0.32***	-0.24*	-0.18	-0.11	-0.01
AWD			0.37***			0.70
High position			0.19			0.09
Legroom			0.42*			2.36
Third row			-0.31***			-0.09
<b>Homophily effects</b>						
Price difference	-1.71***	-1.31***	-1.32***	-1.56***	-1.41***	-1.40***
Power difference	-0.30	-0.18	-0.14	0.60	0.79	0.58
Fuel consump. difference	-0.29***	-0.25***	-0.25***	-0.26	-0.21	-0.15
AIC	122971	122731	122678	1872	1863	1864

\*\*\*p < 0.001, \*\*p < 0.01, \*p < 0.05

<sup>1</sup> Model A-only consider regular attributes and homophily effects

<sup>2</sup> Model B-consider regular attributes, homophily effects and network effects

<sup>3</sup> Model C-consider all attributes

When examining the results from the STERGM formation model of Model C, the coefficients of two network structure effects are significant. The positive sign of “closure effect” indicates that a new competition is more likely to form between two vehicles if they shared the same set of competitors previously. The negative sign of “centralization effect” denotes that new competitive links are more likely to form between vehicles that have been in competition with many other car models already. Among the regular vehicle attributes, the effect of Korean brand origin is significant. Its negative sign indicates that, compared to domestic vehicles, vehicles from Korea are less likely to form new competitive links from 2013 to 2015. Although the main effects of price, power, FC, and turbo show no significance in the model for the information of competitive links, the homophily effects of price and fuel consumption and the main effects of SUV-relevant

attributes such as AWD, legroom and third row are significant. The coefficient of the price difference is -1.32 which indicates that vehicles tend to form new competitions with those having a smaller difference in price. For example, if the price of a car is twice of another car<sup>2</sup>, the log-odds of their competition will be 1.32 lower than the log-odds of competition between two cars with the same price [13]. In addition, two vehicles with a lower difference in FC are also more likely to create a new competition between them as time goes by. The positive coefficients of AWD and legroom means vehicles with all-wheel drive feature and higher legroom rating score are more likely to establish new competition relations with others. In contrast, the negative coefficient of third row indicates that vehicles equipped without third row are more likely to form new competitions with other competitors than vehicles with third row.

When examining the results from the link dissolution model of Model C, it is important to note that the estimations measure the persistence of competition, *not* disappearance. As shown in Table 2, the estimates of closure effect, price difference and vehicles from American brands significantly influence the preservation of existing competitions. The interpretation of the dissolution model is somewhat similar to the formation model except that one is for the formation of new competitions while the other is for the old/existing competitions. For example, the estimate of closure effect being 0.40 indicates that existing competitions are more likely to persist over time if two vehicles share the same set of competitors.

To better understand the differences between STERGM and ERGM, we compare the results of the two models in Table 3. It is noted that the statistical significance of the attributes in ERGM is different from one year to another. In many cases, it is also different from the significance obtained

---

<sup>2</sup> Noted that log2 transformation was used to deal with the attribute values. For example, if we have three cars A, B, and C. Their original price is  $p_1$ ,  $p_2$  and  $p_3$ , separately. Let  $p_1 = p_2$  and  $p_3 = 2 * p_1$ , then the log-odds of existing a competition relation between A and B will be  $\theta \times 0 = 0$ . But the log-odds of existing a competition relation between A and C will be  $\theta \times |\log_2(p_1) - \log_2(p_3)| = \theta = -1.32$ .



from STERGM. These differences can be explained by the fact that ERGM is a static network modeling approach based on single-year data, assuming there are no pre-existing network relations (i.e., no competition at all), whereas STERGM is a dynamic network modeling approach focused on detecting the **changing pattern** that best describes the formation and dissolution of competitions conditional on the pre-existing competitions.

The differences in the model coefficients of ERGM over time (2013-2015 as shown in Table 3) imply the change of customer preferences from one year to another. For instance, the coefficient for fuel consumption difference is insignificant in 2013 but becomes significant in both 2014 and 2015 with a value of -0.20 and -0.23, respectively. On the other hand, the significant coefficient (-0.25) of fuel consumption difference in the STERGM formation model indicates that based on the three-year data, lower fuel consumption difference has a positive influence on forming new competitions. Therefore, to some extent, the STERGM results are able to capture the change of ERGM results. It is also found that the significance of some attributes is different between ERGM and STERGM. For example, seat position is shown to be significant in 2014's static ERGM modeling but insignificant in influencing the formation of new competitions or preservation of old competitions over time compared to other main attributes.

**Table 3. Comparing results of ERGM versus STERGM for competition networks from 2013 to 2015**

	ERGM			STERGM	
	2013	2014	2015	Formation	Dissolution
<b>Network effects</b>					
Edges	7.04*	-2.18	2.40	-1.91	-2.08
Closure	0.77***	1.00***	0.87***	0.85***	0.40**
Centralization	-0.49	0.23	-0.84*	-0.66*	0.50
<b>Main attributes effects</b>					
Price	-0.29*	-0.18	-0.24*	-0.12	-0.36

Power	-0.11	-0.12	-0.01	-0.10	-0.33
Fuel consump.	0.03	0.07*	0.10*	0.05	0.16
Turbo	-0.18	0.09	0.07	0.11	-0.11
Origin (US)	0.13	0.13	-0.14	-0.10	0.65*
Origin (EU)	0.54***	0.33**	0.13	0.03	0.68
Origin (JP)	0.54***	0.23*	-0.09	-0.09	0.06
Origin (KR)	0.50***	-0.08	-0.22	-0.24*	-0.01
AWD	0.53***	0.34***	0.40***	0.37***	0.70
High position	-0.01	0.56*	-0.08	0.19	0.09
Legroom	0.22	0.37	0.36	0.42*	2.36
Third row	-0.25*	-0.17	-0.36***	-0.31***	-0.09
<b>Homophily effects</b>					
Price difference	-1.40***	-1.15***	-1.57***	-1.32***	-1.40***
Power difference	-0.58*	-0.44*	0.35	-0.14	0.58
Fuel consump. difference	-0.10	-0.20***	-0.23***	-0.25***	-0.15
AIC	64672	61255	63363	122678	1864

### 3.5. Results evaluation

To evaluate the results in Section 3.4, link-level evaluation and the goodness-of-fit evaluation analysis at the network level are performed.

#### 3.5.1. Link-level evaluation

Link-level evaluation compares the log-odds of those hypothetical links (i.e., those links do not exist in observed networks and are solely for testing purposes) to the log-odds of the real links in observed networks. In general, newly formed competitions in observed networks (i.e., real links) are expected to have higher log-odds than those hypothetical links. The vehicles represented by green nodes in Fig. 8 are selected for the link-level verification. Table 4 provides the value of significant variables in Model C for these vehicles.

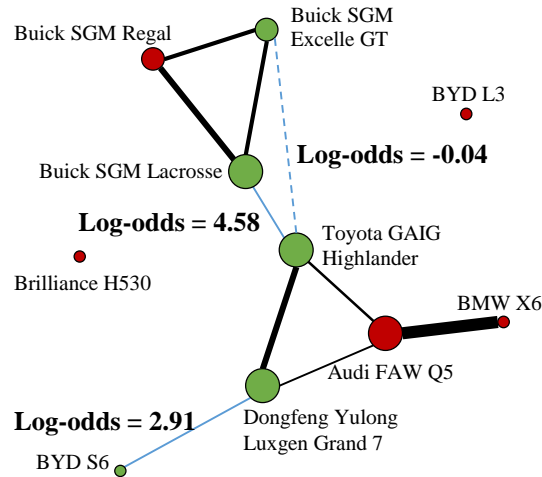
**Table 4. The value of vehicle attributes for selected vehicles**

Model name	Price	FC	Make origin	AWD	Legroom	Third row
Buick SGM Excelle GT	16.69	8.17	American	0	3.09	0
Buick SGM Lacrosse	17.80	11.08	American	0	3.11	0
BYD S6	16.47	9.65	Chinese	0	3.13	0
Dongfeng Yulong luxury grand 7	17.83	11.72	Chinese	0.5	3.19	0
Toyota GAIG Highlander	18.05	11.23	Japanese	0.5	3.19	1

By inserting the estimated coefficients obtained from Table 2 into Eqn. (2), the log-odds of a link forming conditional on the rest of network can be calculated:

$$\text{Logit } Pr(Y_{ij}^+ = 1) = \boldsymbol{\theta}^T \cdot \boldsymbol{\delta}_{ij}^+(y) = 0.85 \times \delta_{GWESP} - 0.66 \times \delta_{GWD} - 0.24 \times \delta_{Origin(KR)} - 1.32 \times \delta_{Price \text{ diff.}} - 0.25 \times \delta_{FC \text{ diff.}} + 0.37 \times \delta_{AWD} + 0.42 \times \delta_{Legroom} - 0.31 \times \delta_{Thirddrow}, \quad (11)$$

Plugging the attribute values from Table 4 into Eqn. (11), we calculate the log-odds of real links compared with a hypothetical link to verify the accuracy of our model; results are illustrated in Fig. 9, using the same set of vehicles shown in Fig. 8. It is observed that the real links (e.g., the competition link between Highlander and Buick SGM Lacrosse) reach higher log-odds than the hypothetical links (e.g., the dash link between Highlander and Buick SGM Excelle GT) in 2015. This indicates that the STERGM results successfully capture the influence of exogenous variables (e.g., fuel consumption, price difference, power difference, fuel consumption difference, and make origin), and endogenous variables (e.g., centralization and closure effect) on the formation of vehicle competitions shown in Fig. 8.

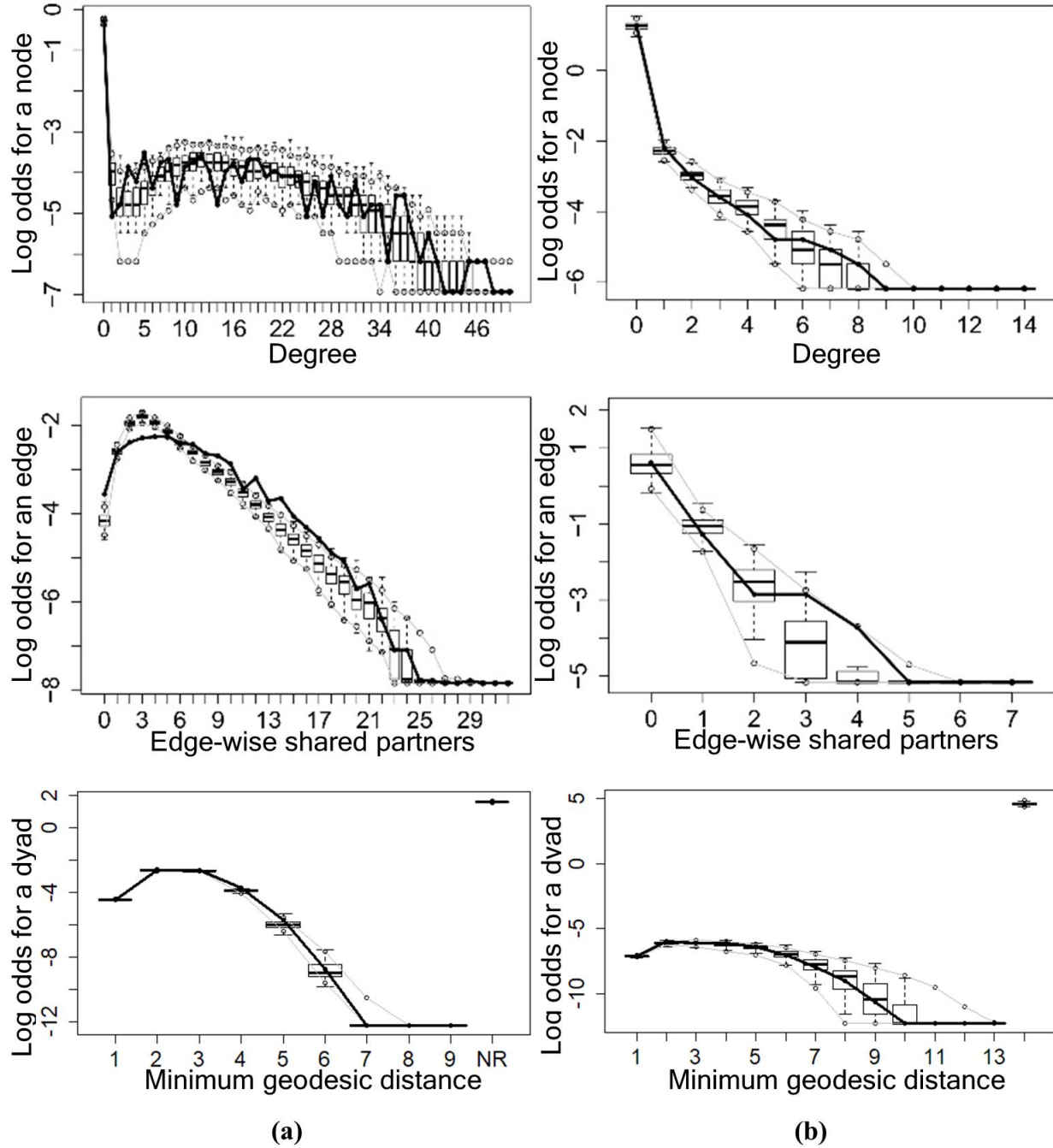


**FIGURE 9: The log-odds results of two newly formed links (solid blue lines) and a hypothetical link (dash blue line) in the competition network from 2014 to 2015.**

### 3.5.2. Network-level evaluation

The goodness-of-fit analysis at the network level evaluates the model by comparing the simulated networks from the estimated models with observed networks in terms of the distributions of certain endogenous variables such as the degree of nodes and GWESP as well as exogenous statistics. In this case study, degree, edgewise shared partner and geodesic distance were selected to evaluate whether simulated networks will reproduce the statistics in an observed network. We use competition network in 2014 as the target data for STERGM simulations. Fig. 10 provides the results of 1000 simulated competition network of 2014 with STERGM (using competition network in 2013 as the starting network) for examining the explanatory variables. The vertical axis in each plot represents the logit (log-odds) of the relative frequency, the solid line represents the statistics for the observed network, the boxplots indicate the median and interquartile range of the simulated networks, and the light-grey lines represent the range in which 95% of simulated observations fall.

We see that most observed value lies in the 95% range of simulated observations which indicates that STERGM performs relatively well in both the formation model and the dissolution model.



**FIGURE 10: Goodness-of-fit plots of STERGM using competition network in 2014 as the target data. (a) Formation model. (b) Dissolution model.**

#### 4. PREDICTABILITY AND APPLICATION IN VEHICLE DESIGN

In this section, we first compare the performance of STERGM (dynamic model) vs. ERGM (static model) in predicting future product competitions. We then present how the proposed approach can be used to inform vehicle design decisions to make a car model more competitive in the market.

##### 4.1. Predictability of dynamic model vs. static model

Compared to ERGM, STERGM is capable of capturing the influence of existing network structures on the formation of future networks. In addition, STERGM is capable of modeling the formation of new links and the dissolution of old links separately. Therefore, STERGM has a higher predictive power than ERGM.

In the following demonstration, for STERGM, we use the models obtained with the datasets from 2013 to 2015 to predict the vehicle competition relations in the 2016 market. For ERGM, we use the model obtained with the 2015 dataset for prediction. The structural zeros method was utilized in STERGM to handle the emerging or delisted vehicle models from 2015 to 2016. In our dataset, 249 car models were observed over the two years, in which 119 models are observed in 2015 and 202 models are in 2016. The change of vehicle competition from 2015 to 2016 is illustrated in Table 5. We see that 17% (76/450) of the competitions in 2015 is preserved in 2016. In both ERGM and STERGM, the values of new vehicle attributes in 2016 and the existing network effects (market structures) in 2015 are utilized for prediction. Compared to other threshold curves, Precision-Recall (PR) curve has been demonstrated by Saito and Rehmsmeier [49] to be especially informative for an imbalanced dataset in which the number of negatives (or positives) outweighs the number of positives (or negatives) significantly, which is the case we face (most cells of the adjacency matrix are zeros since no links exist there). Therefore, we use it to measure the capability of ERGM and STERGM in predicting the 2016 vehicle association (competition) network. In the

PR curve, *precision* is the fraction of true predictions among all predictions, *recall* is the fraction of true predictions among all observations, and area under the PR curve (AUC) indicates predictive performance [49] (larger is better).

**Table 5. The market competition changes from 2015 to 2016**

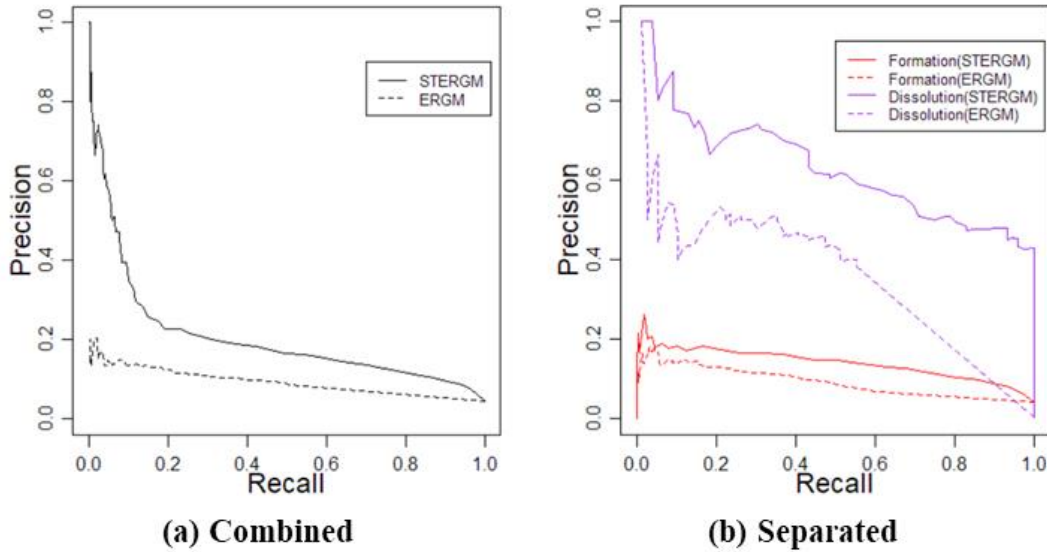
	Competition in 2016	No competitions in 2016	Total
<b>Competition in 2015</b>	76 pairs of vehicles	374 pairs of vehicles	450 pairs of competitions
<b>No competitions in 2015</b>	856 pairs of vehicles	49734 pairs of vehicles	50590 pairs of vehicles
<b>Total</b>	932 pairs of competitions	50108 pairs of vehicles	51040 possible competitions

Figure 11 shows the PR curve with the decision threshold set from 0 to 1 [50] under two situations:

1) Compare the predicted networks with the observed network; 2) Compare the predicted formation network and dissolution network with observed formation network and dissolution network separately. When using STERGM and ERGM, formation and dissolution network are derived from the differences in the original 2015 and predicted 2016 networks. Fig. 11(a) shows that STERGM performs better than ERGM as it has a larger AUC. Fig. 11(b), while consistent with the results from Fig. 11(a), provides additional insights into the network formation and dissolution processes. Specifically, from Fig. 11(b) we observe that:

1) STERGM is better than ERGM in both network formation prediction as well as network preservation prediction. 2) The prediction for preserved links outperforms the prediction for newly formed links in almost all range of the PR curve. This makes sense because the DOF (degree of freedom) of dissolution network (450 pairs) is much lower than formation network (50590 pairs). 3) The gap between STERGM and ERGM for preserved links is much larger than that for formed

links, which indicates the advantage of STERGM for considering the influence of existing market structures on future competitions.



**FIGURE 11: The Precision-Recall curves of STERGM and ERGM.**

In summary, Fig. 11 is consistent with our expectation that STERGM has a higher predictive capability than ERGM in predicting future networks. The latter offers static modeling that assumes there is no preexisting links in the network.

#### 4.2. Application in vehicle design

Numerous studies have identified the importance of enhancing the competitiveness of product through design and technology innovation in improving the performance of firms and national economics [51,52]. In this subsection, we demonstrate how the developed approach can examine the influence of engineering attributes on the competitiveness of a product. These insights can potentially support engineering design and help inform design decisions, e.g., whether to upgrade the design features of an existing vehicle model and by how much, in order to make it more competitive in the market. Here we take the crossover SUV model “Ford Changan Edge” (Ford



Edge) as an example, to study how the adjustment of Ford Edge's engineering attributes will affect its competitiveness in China's auto market.

Table 6 shows the change of vehicle attributes of Ford Edge in the Chinese market from 2015 to 2016. Attributes such as turbo, make origin and AWD stayed the same, price and customers' ratings for high position and legroom increased, and others such as power, fuel consumption and the third row decreased. This change indicates that the upgrade of Ford Edge focused on improving the driving comfort and fuel economy in these two years. We can also observe an increasing price from 2015 to 2016.

**Table 6. The attributes of Ford Changan Edge in 2015 and 2016**

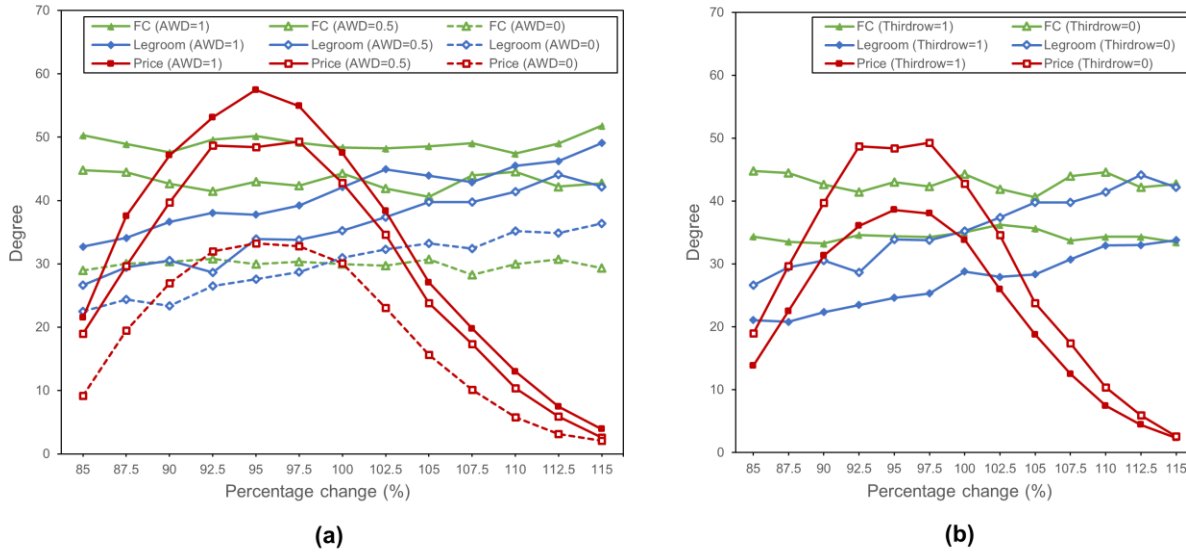
Attributes	2015	2016	Attributes	2015	2016
Price (log2)	18.18	18.23	AWD	0.5	0.5
Power (log2)	7.919	7.917	High position	2.99	3.21
FC (L/100 km)	11.80	11.67	Legroom	3.06	3.31
Turbo	1	1	Third row	1	0
Origin	US	US			

Given the above information, we predict the competitiveness of the modified Ford Edge using the STERGM model. The competitiveness of a vehicle model is measured by the degree centrality metric, which is defined as the number of links connected upon a node (i.e., the number of competitions that a vehicle has) [53]. Therefore, those vehicles with more competitors (i.e., they are co-considered with more other vehicles) are considered as more competitive in the market. Based on the significance of dynamic modeling results obtained in Section 3.4, price, fuel consumption, legroom, AWD and "third row" are selected as changing design attributes. We assume the values of the first three continuous attributes can be changed up to +15% and -15% of their original values, and for categorical attributes, AWD can be changed to 0, 0.5 or 1 and "third row" to 0 or 1. AWD=0.5 stands for a car model with both AWD or non-AWD versions. These

changed attributes are taken as inputs to the STERGM model and 100 predicted competition networks in 2016 are generated and the average value is used to assess the competitiveness of Ford Edge among the 100 predicted networks. Note our study only considers one variable change at a time and no dynamic (game) effects among competitors are considered.

Fig. 12 shows the average trend of vehicle competitiveness with fixed AWD or third row. We can observe that increasing legroom rating can generally make Ford Edge to be more competitive in 2016, which is consistent with dynamic modeling results in Section 3.4. In addition, Fig. 12 indicates that, in most cases, 95% of the original price (17.32 after log2 transformation) can achieve the maximum competitiveness of Ford Edge. This result implies that product competitiveness does not always change monotonously with the price. In order to increase the competitiveness of vehicles, auto companies should improve their products from multiple aspects rather than relying on lowering price exclusively. Different from price and legroom, we cannot see obvious change of vehicle competitiveness when fuel consumption varies. Therefore, fuel economy does appear to be a crucial influencing factor for improving Ford Edge's competitiveness in 2016 China's auto market. The impact of varying AWD and "third row" is shown in Fig. 12(a) and 12(b). In summary, Fig. 12 implies that reasonable price (approximately 95% of the original

value), larger seat space, better traction and control rather than more seats or better fuel economy are more likely to increase the competitiveness of Ford Edge.



**FIGURE 12: The change of predicted degree for Ford Edge with varied price, FC and legroom and fixed original AWD or third row value. (a) Fixed third row = 0; (b) Fixed AWD = 0.5.**

## 5. CONCLUSION

The major contribution of this study is the development of a data-driven dynamic network analysis approach to modeling the evolution of product competition relations. Even though a network-based approach was previously adopted in modeling vehicle competitions to overcome the limitations of DCA [30,54], this is the first attempt to systematically analyze and model product competitions based on longitudinal market data and dynamic network analysis. The structural zeros method is leveraged to tackle the issue of varying product sets from year to year in modeling dynamic competitions. Different from our previous study on multi-year analysis using cross-sectional network data, this research provides insights into the factors (such as product attributes, homophily effects, and network structure effects) that drive changes of product competitions.

Our proposed STERGM approach has the advantage of modeling the impact of endogenous variables as well as exogenous variables on the formation and dissolution of product competitions separately. A three-year customer survey from China's auto market was utilized and three crossover SUV-oriented competition networks were constructed to illustrate the benefits. By utilizing the structural zeros method, we addressed the challenge of modeling network dynamics when the nodes vary from year to year. We observe a significant improvement of model fit after the network structural effects are introduced into the dynamic model. Our verification at both the link level and the network level further demonstrates the model fit.

Our work also illustrates the difference between the static ERGM and the dynamic STERGM. In summary, ERGM is a static network modeling approach assuming there are no pre-existing network relations (i.e., no pre-existing competition at all), whereas STERGM is a dynamic network modeling approach focused on detecting the changing pattern that best describes the formation and dissolution of competitions conditional on the pre-existing competitions.

Our study demonstrates the benefits of using dynamic STERGM over static ERGM for predicting the impact of product design change on a competition network structure. Specifically, the precision-recall test shows a higher predictive power of the STERGM compared to ERGM. Our work also demonstrates the capability of the developed approach to examine the influence of engineering attributes on the competitiveness of products. Therefore, the dynamic network modeling approach is more effective to support engineering design decisions and companies' strategic decision making when improving existing products or releasing new products.

It should be noted that the scope of this work is limited to studying the impact of engineering design attributes on the dynamic evolution of a competition network. In reality, other factors such as positioning, marketing, distribution and retail channels may have an impact. Even though these

factors are inexplicitly captured in the network structure effects and random errors in our current ERGM modeling, explicit modeling of these factors as attributes of network nodes is more desirable when the associated data is available. In addition, since our approach utilized a projection from two-mode network (bipartite graph) to a one-mode network (unidimensional network), the model may somewhat overestimate the triadic closure effect in the network [55]. Nevertheless, our network is not just a direct projection from customer-product system but using lift function as a criterion to determine whether there is a link between two cars. Based on our dataset, it is difficult to say whether GWESP used for triadic closure will overestimate the closure effect of the competition network. In future work, we will conduct further studies to understand this issue. Besides, our future work will also focus on examining the use of the developed approach for different types of network, such as choice network, weighted network and multidimensional network.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support from NSF-CMMI-1436658, Ford-Northwestern Alliance Project, and the grant from China Scholarship Council (No. 201706030108) and MOST Project (No. 2018YFB1700802) for supporting Jian Xie's predoctoral study at Northwestern University. Editorial support provided by Faez Ahmed is greatly appreciated.

## REFERENCES

- [1] Christensen, C., 1997, "Patterns in the Evolution of Product Competition," *Eur. Manag. J.*, **15**(2), pp. 117–127.
- [2] Bloomberg News, 2018, "American Automakers Losing Ground in a Shrinking Chinese Market," *Bloomberg.com*.
- [3] Moss, T., and Colias, M., 2019, "Ford's Shrinking China Business Is Hurting Its Global Ambitions," *Wall Str. J.* [Online]. Available: <https://www.wsj.com/articles/fords-slowng-china-business-is-creating-roadblocks-for-its-global-ambitions-11563787801>.

- [4] Automotive News, 2018, “Ford’s China Sales Fall 32% July on Lack of Fresh Models” [Online]. Available: <https://www.autonews.com/article/20180810/GLOBAL/180819982/ford-s-china-sales-fall-32-july-on-lack-of-fresh-models>.
- [5] Turrentine, T. S., and Kurani, K. S., 2007, “Car Buyers and Fuel Economy?,” *Energy Policy*, **35**(2), pp. 1213–1223.
- [6] Heffner, R., Kurani, K. S., and Turrentine, T. S., 2007, “Symbolism in Early Markets for Hybrid Electric Vehicles,” Tech. Rep. No. UCD-ITS-RR-07-01, Inst. Transp. Stud. Univ. California, Davis, Davis, CA.
- [7] Talay, M. B., Calantone, R. J., and Voorhees, C. M., 2014, “Coevolutionary Dynamics of Automotive Competition: Product Innovation, Change, and Marketplace Survival,” *J. Prod. Innov. Manag.*, **31**(1), pp. 61–78.
- [8] Roberts, P. W., 1999, “Product Innovation, Product–Market Competition and Persistent Profitability in the US Pharmaceutical Industry,” *Strateg. Manag. J.*, **20**(7), pp. 655–670.
- [9] Gandai, N., 2001, “The Dynamics of Competition in the Internet Search Engine Market,” *Int. J. Ind. Organ.*, **19**(7), pp. 1103–1117.
- [10] Chisholm, D. C., and Norman, G., 2012, “Market Access and Competition in Product Lines,” *Int. J. Ind. Organ.*, **30**(5), pp. 429–435.
- [11] Medhin, N. G., and Wan, W., 2009, “Multi-New Product Competition in Duopoly: A Differential Game Analysis,” *Dyn. Syst. Appl.*, **18**(2), p. 161.
- [12] Wang, M., Sha, Z., Huang, Y., Contractor, N., Fu, Y., and Chen, W., 2018, “Predicting Products’ Co-Considerations and Market Competitions for Technology-Driven Product Design: A Network-Based Approach,” *Des. Sci. J.*, **4**.
- [13] Sha, Z., Huang, Y., Fu, S., Wang, M., Fu, Y., Contractor, N., and Chen, W., 2018, “A Network-Based Approach to Modeling and Predicting Product Co-Consideration Relations,” *Complexity*, **2018**.
- [14] Robins, G., Pattison, P., Kalish, Y., and Lusher, D., 2007, “An Introduction to Exponential Random Graph (P\*) Models for Social Networks,” *Soc. Networks*, **29**(2), pp. 173–191.
- [15] Wang, M., Chen, W., Huang, Y., Contractor, N. S., and Fu, Y., 2016, “Modeling Customer Preferences Using Multidimensional Network Analysis in Engineering Design,” *Des. Sci.*, **2**.
- [16] Snijders, T. A. B., Pattison, P. E., Robins, G. L., and Handcock, M. S., 2006, “New Specifications for Exponential Random Graph Models,” *Sociol. Methodol.*, **36**(1), pp. 99–153.
- [17] Wang, M., Sha, Z., Huang, Y., Contractor, N., Fu, Y., and Chen, W., 2016, “Forecasting Technological Impacts on Customers’ Co-Consideration Behaviors: A Data-Driven Network Analysis Approach,” *ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Charlotte, NC, USA, August 21–24, 2016, p. V02AT03A040--V02AT03A040.
- [18] Wang, M., Chen, W., Fu, Y., and Yang, Y., 2015, “Analyzing and Predicting Heterogeneous Customer Preferences in China’s Auto Market Using Choice Modeling and Network Analysis,” *SAE Int. J. Mater. Manuf.*, **8**(2015-01–0468), pp. 668–677.
- [19] Fu, J. S., Sha, Z., Huang, Y., Wang, M., Fu, Y., and Chen, W., 2017, “Modeling Customer Choice Preferences in Engineering Design Using Bipartite Network Analysis,” *Proceedings of the ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Cleveland, Ohio.

- [20] Hanneke, S., Fu, W., Xing, E. P., and others, 2010, “Discrete Temporal Models of Social Networks,” *Electron. J. Stat.*, **4**, pp. 585–605.
- [21] Krivitsky, P. N., and Handcock, M. S., 2014, “A Separable Model for Dynamic Networks,” *J. R. Stat. Soc. Ser. B (Statistical Methodol.)*, **76**(1), pp. 29–46.
- [22] Watts, D. J., and Strogatz, S. H., 1998, “Collective Dynamics of ‘small-World’ networks,” *Nature*, **393**(6684), p. 440.
- [23] Barabási, A.-L., and Albert, R., 1999, “Emergence of Scaling in Random Networks,” *Science* (80-. ), **286**(5439), pp. 509–512.
- [24] Xu, K. S., and Hero, A. O., 2013, “Dynamic Stochastic Blockmodels: Statistical Models for Time-Evolving Networks,” *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, Washington, DC, USA, April 2-5, 2013, pp. 201–210.
- [25] Block, P., Koskinen, J., Hollway, J., Steglich, C., and Stadtfeld, C., 2018, “Change We Can Believe in: Comparing Longitudinal Network Models on Consistency, Interpretability and Predictive Power,” *Soc. Networks*, **52**, pp. 180–191.
- [26] Snijders, T. A. B., de Bunt, G. G., and Steglich, C. E. G., 2010, “Introduction to Stochastic Actor-Based Models for Network Dynamics,” *Soc. Networks*, **32**(1), pp. 44–60.
- [27] Leifeld, P., and Cranmer, S. J., 2015, “A Theoretical and Empirical Comparison of the Temporal Exponential Random Graph Model and the Stochastic Actor-Oriented Model,” *arXiv Prepr. arXiv1506.06696*.
- [28] Mousavi, R., and Gu, B., 2015, “The Effects of Homophily in Twitter Communication Network of US House Representatives: A Dynamic Network Study,” Available SSRN <https://ssrn.com/abstract=2666052> or <http://dx.doi.org/10.2139/ssrn.2666052>.
- [29] Lebacher, M., Thurner, P. W., and Kauermann, G., 2018, “International Arms Trade: A Dynamic Separable Network Model With Heterogeneity Components,” *arXiv Prepr. arXiv1803.02707*.
- [30] Wang, M., and Chen, W., 2015, “A Data-Driven Network Analysis Approach to Predicting Customer Choice Sets for Choice Modeling in Engineering Design,” *J. Mech. Des.*, **137**(7), p. 71410.
- [31] Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M., 2008, “Ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks,” *J. Stat. Softw.*, **24**(3), p. nihpa54860.
- [32] Lusher, D., Koskinen, J., and Robins, G., 2012, *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*, Cambridge University Press.
- [33] Han, J.-D. J., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., Dupuy, D., Walhout, A. J. M., Cusick, M. E., Roth, F. P., and others, 2004, “Evidence for Dynamically Organized Modularity in the Yeast Protein--Protein Interaction Network,” *Nature*, **430**(6995), p. 88.
- [34] Conaldi, G., and Lomi, A., 2013, “The Dual Network Structure of Organizational Problem Solving: A Case Study on Open Source Software Development,” *Soc. Networks*, **35**(2), pp. 237–250.
- [35] Tsamardinos, I., and Borboudakis, G., 2010, “Permutation Testing Improves Bayesian Network Learning,” *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 322–337.
- [36] Hu, M.-C., Pavlicova, M., and Nunes, E. V., 2011, “Zero-Inflated and Hurdle Models of Count Data with Extra Zeros: Examples from an HIV-Risk Reduction Intervention Trial,” *Am. J. Drug Alcohol Abuse*, **37**(5), pp. 367–375.

- [37] Blanchini, F., Franco, E., and Giordano, G., 2013, “Structured-LMI Conditions for Stabilizing Network-Decentralized Control,” *52nd IEEE Conference on Decision and Control*, pp. 6880–6885.
- [38] Robins, G., Pattison, P., and Woolcock, J., 2004, “Missing Data in Networks: Exponential Random Graph (P\*) Models for Networks with Non-Respondents,” *Soc. Networks*, **26**(3), pp. 257–283.
- [39] Snijders, T. A. B., 1991, “Enumeration and Simulation Methods for 0--1 Matrices with given Marginals,” *Psychometrika*, **56**(3), pp. 397–417.
- [40] Sha, Z., Bi, Y., Wang, M., Stathopoulos, A. Contractor, N., Fu, Y., and Chen, W., 2019, “Comparing Utility-Based and Network-Based Approaches in Estimating Customer Preferences for Engineering Design,” *The 21st International Conference on Engineering Design*, Delft, The Netherlands.
- [41] Tan, P.-N., Kumar, V., and Srivastava, J., 2004, “Selecting the Right Objective Measure for Association Analysis,” *Inf. Syst.*, **29**(4), pp. 293–313.
- [42] Wang, M., Sha, Z., Huang, Y., Contractor, N., Fu, Y., and Chen, W., 2018, “Predicting Product Co-Consideration and Market Competitions for Technology-Driven Product Design: A Network-Based Approach,” *Des. Sci.*, **4**.
- [43] Sha, Z., Wang, M., Huang, Y., Contractor, N., Fu, Y., and Chen, W., 2017, “Modeling Product Co-Consideration Relations: A Comparative Study of Two Network Models,” *Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 6: Design Information and Knowledge, Vancouver, Canada, 21-25.08. 2017*.
- [44] Stevens, S. S., and others, 1946, “On the Theory of Scales of Measurement.”
- [45] Kostoulas, A., 2013, “On Likert Scales, Ordinal Data and Mean Values” [Online]. Available: <https://achilleaskostoulas.com/2013/02/13/on-likert-scales-ordinal-data-and-mean-values/>.
- [46] Krivitsky, P. N., Handcock, M., and others, 2018, “Tergm: Fit, Simulate and Diagnose Models for Network Evolution Based on Exponential-Family Random Graph Models,” *Statnet Proj.* (<http://www.statnet.org>). R Packag. version, **3**(2).
- [47] Hunter, D. R., 2007, “Curved Exponential Family Models for Social Networks,” *Soc. Networks*, **29**(2), pp. 216–230.
- [48] Akaike, H., 1998, “Information Theory and an Extension of the Maximum Likelihood Principle,” *Selected Papers of Hirotugu Akaike*, Springer, pp. 199–213.
- [49] Saito, T., and Rehmsmeier, M., 2015, “The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets,” *PLoS One*, **10**(3), p. e0118432.
- [50] Flach, P., and Kull, M., 2015, “Precision-Recall-Gain Curves: PR Analysis Done Right,” *Advances in Neural Information Processing Systems*, pp. 838–846.
- [51] Roy, R., and Riedel, J. C., 1997, “Design and Innovation in Successful Product Competition,” *Technovation*, **17**(10), pp. 537–594.
- [52] Hsu, Y., 2011, “Design Innovation and Marketing Strategy in Successful Product Competition,” *J. Bus. Ind. Mark.*, **26**(4), pp. 223–236.
- [53] Opsahl, T., Agneessens, F., and Skvoretz, J., 2010, “Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths,” *Soc. Networks*, **32**(3), pp. 245–251.
- [54] Wang, M., Chen, W., Huang, Y., Contractor, N. S., and Fu, Y., 2015, “A Multidimensional Network Approach for Modeling Customer-Product Relations in Engineering Design,”



*ASME 2015 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, Boston, MA.

- [55] Opsahl, T., 2013, "Triadic Closure in Two-Mode Networks: Redefining the Global and Local Clustering Coefficients," *Soc. Networks*, **35**(2), pp. 159–167.